Identification and characterisation of HNF1B as a pro-tumorigenic lineage factor in clear cell renal cell carcinoma

Emma Kay Richardson

Homerton College, Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

October 2018

Emma Kay Richardson

Identification and characterisation of HNF1B as a pro-tumorigenic lineage factor in clear cell renal cell carcinoma

ABSTRACT:

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cancer. ccRCCs exhibit considerable genetic intratumour heterogeneity, and current therapies lack long-term efficacy due to drug resistance. Novel therapeutic targets are thus required to combat this disease.

The most effective targets are likely to be factors essential for maintaining ccRCC cell phenotypes and survival across all tumour subclones. Transcription factors (TFs) are appealing candidates as they are key determinants of cellular phenotypes, including in cancer. There is growing evidence to suggest that lineage-specific TFs often play important roles in tumourigenesis and tumour progression. ccRCC phenotypes may also be fundamentally defined by tissue-specific core transcriptional networks. Thus, the aim of my PhD thesis was to identify novel TF dependencies in ccRCC, and characterise the mechanisms by which these essential TFs support these tumours.

I performed a CRISPR/Cas9-based genetic depletion screen *in vitro* with an sgRNA library targeting 50 TFs that are highly expressed in RCC. This highlighted HNF1B as a potential tumour dependency in ccRCC. I demonstrated that both HNF1B CRISPR knockout and CRISPRi knockdown conferred a strong selective disadvantage *in vitro* that was rescued by HNF1B restoration, suggesting that it was a specific effect of HNF1B loss. I also showed that HNF1B knockout impaired tumour initiation and growth *in vivo*.

I showed that HNF1B knockout led to a downregulation of MYC expression, along with a cell cycle defect in G1/S progression and an increase in cell death. HNF1B ChIP-seq in ccRCC cell lines revealed HNF1B binding peaks near the *MYC* gene, suggesting that HNF1B might directly regulate MYC in ccRCC. I subsequently found that MYC CRISPRi knockdown phenocopied the striking selective disadvantage associated with HNF1B depletion. Taken together, my data suggests that HNF1B-mediated regulation of MYC might be a key mechanism underlying the importance of HNF1B in ccRCC.

Preface

The work described in this dissertation was carried out at the MRC Cancer Unit, Cambridge, United Kingdom during the period of October 2014 to October 2018. This work was carried out under the supervision of Dr. Sakari Vanharanta. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University of similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. Finally, this dissertation does not exceed the prescribed 60,000 word limit for the Biology Degree Committee.

Acknowledgements

I would like to thank my supervisor, Sakari Vanharanta, for giving me the opportunity to undertake a PhD in his lab, and for his mentorship, teaching, and support. I would also like to thank the other members of the Vanharanta lab, Paulo Rodrigues, Saroor Patel, Saiful Syafruddin (Fendi), Erika Vojtasova, Nazhif Zaini, Ludo Wesolowski, and Annabel Zelcewski for their support and assistance over the years. In particular I would like to express my gratitude to Paulo and Saroor for their mentorship and teaching.

From an experimental standpoint, I would like to thank Saroor Patel for her help with the CRISPR/Cas9 genetic depletion screen, Paulo Rodrigues and Saiful Syafruddin (Fendi) for their help with the ChIP-seq and RNA-seq experiments (including the preparation of the libraries), and Fendi for his help with several of the competition proliferation assays. Many thanks as well to Sakari Vanharanta for all the guidance throughout my experiments, and for helping with the Illumina sequencing data analysis for the screen, ChIP-seq and RNA-seq. I would also like to thank Shamith Samarajiwa and Dora Bihary for their substantial help with the RNA-seq data analysis. I am also extremely grateful to Alyson Speed for her invaluable assistance with my animal experiments, for patiently teaching me how to overcome my fear of mice and handle them for my experiments, and for making the hours spent in the animal facility a lot more enjoyable.

Special thanks to everyone in the MRC-CU for all their support and experimental advice when I approached them with questions, and for making it such a warm and friendly institute to work in.

Most of all, thank you to my wonderful mother, my amazing sister Sara, the love of my life You Yi, and all my fantastic friends in Cambridge who provided me with tons of incredible support to help get me through the past four years. I could not have done it without you by my side – you are all the brightest, happiest parts of my life. You helped me get through the stress of bereavement and a 6,800 mile long distance relationship on top of the rigorous demands of a PhD, and for that I am eternally grateful. I could not have done this without you guys, and I love you all immensely.

Table of Contents

ABSTRACT:

Clear cell renal cell carcinoma (ccRCC) is the most common subtype of renal cancer. ccRCCs exhibit considerable genetic intratumour heterogeneity, and current therapies lack long-term efficacy due to drug resistance. Novel therapeutic targets are thus required to combat this disease.

The most effective targets are likely to be factors essential for maintaining ccRCC cell phenotypes and survival across all tumour subclones. Transcription factors (TFs) are appealing candidates as they are key determinants of cellular phenotypes, including in cancer. There is growing evidence to suggest that lineage-specific TFs often play important roles in tumourigenesis and tumour progression. ccRCC phenotypes may also be fundamentally defined by tissue-specific core transcriptional networks. Thus, the aim of my PhD thesis was to identify novel TF dependencies in ccRCC, and characterise the mechanisms by which these essential TFs support these tumours.

I performed a CRISPR/Cas9-based genetic depletion screen *in vitro* with an sgRNA library targeting 50 TFs that are highly expressed in RCC. This highlighted HNF1B as a potential tumour dependency in ccRCC. I demonstrated that both HNF1B CRISPR knockout and CRISPRi knockdown conferred a strong selective disadvantage *in vitro* that was rescued by HNF1B restoration, suggesting that it was a specific effect of HNF1B loss. I also showed that HNF1B knockout impaired tumour initiation and growth *in vivo*.

I showed that HNF1B knockout led to a downregulation of MYC expression, along with a cell cycle defect in G1/S progression and an increase in cell death. HNF1B ChIP-seq in ccRCC cell lines revealed HNF1B binding peaks near the *MYC* gene, suggesting that HNF1B might directly regulate MYC in ccRCC. I subsequently found that MYC CRISPRi knockdown phenocopied the striking selective disadvantage associated with HNF1B depletion. Taken together, my data suggests that HNF1B-mediated regulation of MYC might be a key mechanism underlying the importance of HNF1B in ccRCC.

ABBREVIATIONS:

Abbreviation	Description
(e)GFP	(enhanced) green fluorescent protein
APC	Allophycocyanin
ARNT	Aryl Hydrocarbon Receptor Nuclear Translocator (gene name for HIF1β)
B-actin	Beta-actin, used as a protein loading control for Western blots
BCL2	B-Cell CLL/Lymphoma 2; BCL2 apoptosis regulator
BFP	blue fluorescent protein
bp	basepair(s)
BSA	bovine serum albumin
Cas	CRISPR-associated proteins
Cas9	CRISPRassociated protein 9
ccRCC	clear cell renal cell carcinoma
CDH16	Cadherin 16
Cdk	cyclin-dependent kinase
cDNA	complementary DNA
ChIP	chromatin immunoprecipitation
ChRCC	chromophobe renal cell carcinoma
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRi	CRISPR interference
Ctrl	Control
dCas9	nuclease-deactivated Cas9 (i.e. a catalytically dead Cas9 protein)
DNA	deoxyribonucleic acid
DOX / Dox / dox	doxycycline
DSB	double-stranded break
EMT	epithelial-to-mesenchymal transition
EPAS1	Endothelial PAS Domain Protein 1 ***or*** Endothelial PAS Domain-Containing Protein 1 (gene name for HIF2A)
EtOH	ethanol
EV or (-)	empty vector
FACs (sorting)	fluorescence-activated cell sorting
FACs analysis	i.e. flow cytometry analysis
FCS	foetal calf serum
FDR	false discovery rate
gDNA	genomic DNA
GSEA	gene set enrichment analysis
HIF1α	Hypoxia-Inducible Factor 1-Alpha
HIF1β	Hypoxia-Inducible Factor 1-Beta
HIF2α (or HIF2A)	Hypoxia-Inducible Factor 2-Alpha
HNF1A	Hepatocyte Nuclear Factor 1-Alpha, a.k.a HNF1 homeobox A
Hnf1b	mouse homologue of *HNF1B*
HNF1B or *HNF1B*	Hepatocyte Nuclear Factor 1-Beta, a.k.a HNF1 homeobox B
HPA	Human Protein Atlas
HR	homologous recombination
IHC	immunohistochemistry

indels	insertions/deletions
IP	immunoprecipitation
kDa	kilodalton(s)
KRAB / KRAB domain	Krüppel-associated box domain
lncRNA	long non-coding RNA
MODY5	maturity-onset diabetes of the young, type 5
MOI	multiplicity of infection
mRNA	messenger RNA
MYC	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog; proto-oncogene, bHLH transcription factor
NGS	Next Generation Sequencing
NHEJ	non-homologous end joining
NSG	Nod-Scid-γ
NTC	non-targeting control
NTC18	non-targeting control 18, used as the control sgRNA for experiments
P/S	pen-strep
padj	adjusted p-value
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PI	Propidium iodide
PRCC	papillary renal cell carcinoma
PS	phosphatidylserine
PVT1	plasmacytoma variant translocation 1
qPCR	quantitative (real-time) PCR
qRT-PCR	quantitative reverse-transcriptase polymerase chain reaction
RCAD (syndrome)	renal cysts and diabetes (syndrome)
RCC	renal cell carcinoma
RNA	ribonucleic acid
RNAi	RNA interference
RT-PCR	reverse transcriptase PCR
sg2	sgRNA 2, targeting HNF1B
sg8	sgRNA 8, targeting HNF1B
sgRNA	single-guide RNA
shRNA	short hairpin RNA
SPRI beads	Solid Phase Reversible Immobilization beads
TCGA	The Cancer Genome Atlas
TF	transcription factor
TUNEL assay	terminal deoxynucleotidyl transferase dUTP nick end labeling assay
U/T	untransduced, or untreated (as indicated in the text)
Vhl	mouse homologue of *VHL*
VHL or *VHL*	von-Hippel Lindau (a tumour suppressor gene); von Hippel-Lindau tumor suppressor
vhnf1	zebrafish homologue of *HNF1B*

INTRODUCTION:

Renal cell carcinoma

Renal cell carcinoma (RCC) is the most common malignant tumour type of the kidney, accounting for 2% of all cancers and 80-85% of renal cancers. RCC originates in the renal coritcal epithelium, and is a heterogeneous disease classifiable into several subtypes. My project focuses on the most common subtype, clear cell renal cell carcinoma (ccRCC), which comprises 75-85% of RCC tumours and arises from cells in the proximal tubule of the nephron.[1,2]

ccRCC can be highly aggressive; roughly one-third of patients suffer from metastatic ccRCC. Metastases can form at a wide range of secondary sites in the body, most frequently in the lungs, bone and brain, and the formation of secondary lesions can occur years after initial diagnosis and treatment of the disease. The highly metastatic nature of ccRCC is compounded by the difficulty of early diagnosis, as early stage tumours generally exhibit a lack of symptoms and are non-palpable until the disease progresses. While surgery can be curative in the case of localised ccRCC, metastatic disease at the time of diagnosis or disease recurrence after surgery ultimately result in fatality.[2–7]

Furthermore, ccRCC is resistant to both radiation therapy and standard cytotoxic chemotherapy[1,2,5,8–11]. Various other treatment strategies have been attempted[1,2,10–13], including cytokine-based immunotherapies using interleukin-2 (IL-2) and interferon-α (IFNα); however, these have shown limited efficacy in patients with metastatic disease[9,14–20]. Nivolumab, a recently developed fully human immunoglobulin G4 antibody targeting programmed death-1 (PD-1), has shown more promising results in clinical trials[21–23].

Kinase inhibitors, anti-angiogenic drugs and mTOR inhibitors have also proven to be effective[1,2,32–41,24,42–46,25–31], and the use of these agents has become the current standard of care to treat advanced ccRCC. However, these targeted systemic therapies also do not elicit complete responses and have still failed to provide long-term treatment strategies capable of reliably circumventing the problem of tumour drug resistance[47]. As a result, there is frequent tumour relapse, eventual ccRCC disease progression and patient fatality. Thus, there is a need for novel therapeutic strategies and targets in order to combat ccRCC, particularly in its advanced metastatic stages.

(cc)RCC genetics

Most cases of RCC are nonhereditary, although a small proportion of tumours develop as a result of familial RCC syndromes. In the case of ccRCC, tumours can either occur sporadically or in a hereditary manner in association with von-Hippel Lindau (VHL) disease, a cancer syndrome that follows an autosomal dominant pattern of inheritance. VHL disease predisposes the individual to the development of multiple tumours or cysts in various tissues, including hemangioblastomas (blood vessel tumours of the retina and central nervous system), pheochromocytomas (tumours of the medulla in the adrenal gland), tumours in the islet cells of the pancreas, renal cysts and clear cell renal carcinomas in the kidneys. RCC develops in almost 40% of VHL disease patients, and is a major cause of death in these individuals.[1,2,24,25]

VHL disease results from heterozygous loss of the *VHL* tumour suppressor gene (TSG), located on chromosome 3p25[1,2,24,48]. Individuals possess only one wild-type *VHL* allele, and carry inherited germ-line inactivating mutations in the other allele. Subsequent tumour development (including RCC) then arises from loss or somatic inactivation of the remaining wild-type allele, resulting in loss of function of the VHL protein. *VHL* is thus a classic tumour suppressor gene, and conforms to Knudson's two-hit hypothesis[2,24,25].

Similarly, consistent with the Knudson two-hit model, sporadic ccRCC tumours also typically undergo biallelic inactivation of *VHL*, either via somatic mutation or epigenetic silencing (DNA methylation) of both alleles, or via the loss of one allele through loss of heterozygosity at chromosome 3p with concomitant loss of the other allele by inactivating mutation. Germline mutations are rare in sporadic ccRCC; both "hits" to inactivate each *VHL* allele occur somatically.[1,2,24,25,48–50]

Indeed, inactivation of *VHL* is the most frequent genetic lesion associated with the development of sporadic ccRCC[1,2,24,25,48]. TCGA projects and various other sequencing efforts[49–54] have identified the critical underlying mutations in ccRCC tumours. Multi-region sequencing of spatially separated ccRCC samples from both primary tumours and associated metastases revealed that only about a third of mutations detected were ubiquitously present across all tumour regions[50,51]. Most prominent among these were *VHL* mutations and aberrations in chromosome 3p.

Meanwhile, the majority of somatic driver mutations in ccRCC were found to be heterogeneously distributed between different tumour regions or unique to single regions. These

studies by Gerlinger *et al.* have also revealed a branched (as opposed to linear) evolutionary pattern of ccRCC tumour growth, and highlighted *VHL* mutations as critical driver, or "truncal", events in ccRCC evolution.[50,51]

VHL function and the VHL-HIF pathway

VHL binds via its C-terminal α-domain to elongin B and elongin C, and subsequently forms a complex with cullin-2 (Cul2) and RING-box protein 1 (Rbx1, also known as ROC1 or Hrt1). This complex acts as an E3 ubiquitin-ligase with VHL mediating substrate recognition via its N-terminal β-domain, and catalyses the polyubiquitylation of specific proteins, targeting them for degradation by the proteasome.[1,2,24,25,48]

It is by this mechanism that VHL targets two hypoxia-inducible factor (HIF) subunits involved in ccRCC, HIF1α and HIF2α, for proteasomal degradation. HIF is a heterodimeric transcription factor, comprising an α-subunit (HIF1α or HIF2α) and a β-subunit (HIF1β / ARNT). VHL binds and targets HIFα subunits for degradation only under normoxic conditions, via an oxygen-dependent mechanism during which the HIFα subunits must first undergo prolyl hydroxylation to facilitate recognition by the VHL β-domain. HIFα proteins are thus highly unstable in the presence of oxygen, as they are rapidly polyubiquitylated and degraded. Under hypoxic conditions however, HIFα subunits are not prolyl hydroxylated; HIFα is therefore stabilised and can dimerise with HIF1β to form an active HIF heterodimer, which then activates the transcription of various hypoxia-inducible genes.[2,24,25,48]

ccRCC-associated *VHL* mutations tend to affect the function of the VHL α-domain and/or β-domain, thus impairing formation of the E3 ubiquitin ligase complex or compromising HIFα binding[24,55–62]. Consequently, *VHL* inactivation in ccRCC results in HIF1α / HIF2α stabilisation and accumulation. This results in the overexpression of downstream HIF transcriptional target genes, many of which may have roles in promoting tumour survival, including growth factors such as TGF-α (transforming growth-factor-α) and its receptor EGFR (epidermal growth factor receptor), and angiogenic factors such as VEGF (vascular endothelial growth factor) and PDGF-β (platelet-derived growth-factor B chain). Indeed, ccRCC tumours are typically highly vascular, most likely due to the overproduction of angiogenic factors like

VEGF. Furthermore, inhibition of VEGF signalling can suppress RCC tumour growth, and disruption of TGFα/EGFR signalling inhibits RCC cell proliferation both *in vitro* and *in vivo*. This therefore implicates deregulated HIF transcriptional activity and VHL-HIF signalling in ccRCC tumour initiation and maintenance.[1,2,24,25,48,55,63,64]

VHL and HIF in ccRCC

There is considerable evidence in the literature highlighting the VHL-HIF pathway as a critical driver pathway for tumour formation in RCC. Constitutive restoration of wild-type VHL in VHL-/- ccRCC cells is sufficient to suppress tumour growth *in vivo*[65], in a scenario representative of VHL reintroduction at the point of tumour initiation. VHL restoration has also been shown to reduce cell proliferation and promote cell cycle exit and differentiation *in vitro*, in certain contexts[66–70]. This indicates that ccRCC tumours remain dependent on the absence of functional VHL, at least in the early stages during tumour initiation.

Meanwhile, studies using HIF2α mutants defective for VHL binding have demonstrated that downregulation of HIF2α is necessary and sufficient to facilitate pVHL-mediated tumour suppression. HIF2α stabilisation and consequent target gene activation was sufficient to override VHL tumour suppression and promote tumour growth in VHL-restored ccRCC cells *in vivo*. These effects were abrogated if the DNA-binding capacity of HIF2α was impaired, and were thus linked to the ability of HIF2α to activate transcription. Moreover, shRNA knockdown of HIF2α in VHL-/- ccRCC cells has been shown to be sufficient to suppress tumour growth *in vivo*. Collectively, this evidence suggests that HIF2α plays a critical role in ccRCC tumourigenesis. This is further underscored by immunostaining analysis of kidneys from VHL disease patients.[55,63,64,71]

Similar studies have been conducted in the same ccRCC cell lines (which express HIF2α but not HIF1α) using a HIF1α mutant defective for VHL recognition. However, despite this HIF1α mutant being transcriptionally active, it was not sufficient to overcome VHL-mediated tumour suppression in VHL-restored ccRCC cells and thus did not produce the same tumourigenic phenotype *in vivo*. Instead, presence of the non-degradable HIF1α mutant tended to reduce the tumourigenic potential of the cells even further than the presence of wild-type

VHL. This suggests that HIF1α is not as critical for ccRCC tumourigenesis as HIF2α, and does not exhibit oncogenic behaviour (or is simply less oncogenic) in this context.[72]

However, despite the apparent importance of the VHL-HIF pathway in ccRCC tumour initiation and maintenance, VHL and HIF do not seem to be the only factors contributing to ccRCC survival. Although VHL restoration suppresses ccRCC cell proliferation, it does not abrogate it entirely, suggesting that there may be other pathways supporting tumour survival.

Additionally, although *VHL* is expressed in a wide variety of foetal and adult tissues, *VHL* inactivation only promotes tumourigenesis in a limited range of tissue types, namely the kidney and other organs affected in VHL disease. Other types of sporadic tumours, such as those of the breast, colon, lung and prostate, rarely exhibit somatic mutations in *VHL*. Furthermore, different *VHL* mutations give rise to different tumour types in VHL disease, and different levels of ccRCC risk. The reasons for these trends are unclear, but these observations continue to suggest that perturbations in other genes or pathways apart from VHL-HIF may play a role in ccRCC development.[24,25,55,64]

Indeed, there have been no reports to suggest that *VHL* inactivation alone initiates ccRCC development. Loss of *VHL* occurs as an early event in RCC development and biallelic *VHL* inactivation has been detected in ccRCC tumours, but it has also been linked to the development of *premalignant* renal cysts that still overexpress HIF and HIF targets like VEGF. This suggests that the *VHL* tumour suppressor gene may simply act as a "gatekeeper", and that additional mutations in other genes are required for progression to outright ccRCC disease.[24,25,55,63,64]

In keeping with this idea, although homozygous *Vhl* knockout (*Vhl-/-*) has been found to be embryonic lethal in mice, there have been mixed results in the literature in studies of heterozygous *Vhl* knockout mice (*Vhl+/-*): Gnarra *et al*. showed that the mice developed normally with an absence of tumours[73], while a subsequent study by Haase *et al*. using an independent *Vhl+/-* mouse strain only showed the formation of multiple hepatic lesions that exhibited deregulated expression of HIF transcriptional targets[74]. Haase *et al.* also reported the development of similar lesions in the liver upon conditional *Vhl* inactivation.

However, *VHL* loss seems to be the only obligatory genetic event in ccRCC, and no other ubiquitous "truncal" mutations have been identified so far. This might suggest that progression to ccRCC after VHL inactivation requires, for example, epigenetic alterations or changes in the output of transcriptional programmes, rather than additional genetic mutations.

In the case of HIF2α, recent studies[75,76] suggest that ccRCC cell lines and tumours exhibit differential HIF2α dependence. Only a subset of ccRCC cell lines is sensitive to genetic disruption or drug-mediated inhibition of HIF2α *in vitro* and *in vivo*. HIF2α dependence correlates loosely with the levels of endogenous HIF2α and the dependency of HIF target genes on HIF2α in a given cell line. ccRCC tumours that are sensitive to small molecule inhibition of HIF2α have a distinctive gene expression signature and higher basal HIF2α levels, and prolonged drug treatment resulted in the development of resistance. In some cases, resistance to HIF2α inhibition appeared to be related to the acquisition of mutations in p53 or the p53 pathway. This further underscores the concept that pathways other than the VHL-HIF pathway may play an important role in ccRCC.

Using CRISPR/Cas9 to screen for additional supporting factors in ccRCC

With the limited efficacy of current therapies in the clinic, there is a need to identify novel drug targets in ccRCC. To that end, I have adopted a negative genetic screening approach using the CRISPR/Cas9-sgRNA system[77–79].

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)-Cas (CRISPR-associated proteins) system essentially constitutes the adaptive immune system in bacteria and archaea, protecting against invasion by viruses and foreign plasmids[80]. There are three types of CRISPR-Cas system; Type II is the simplest, and involves the cleavage activity of the Cas9 protein. The components of this prokaryotic defence mechanism have subsequently been adapted for applications in genetic engineering.[81–83]

Two components are required: the Cas9 protein and a 20bp targeting sequence incorporated within a chimeric guide RNA structure (called a single-guide RNA, sgRNA). These are typically expressed under the control of a U6 promoter in the same or separate expression plasmids, which can be stably or transiently introduced into mammalian cells. The sgRNA directs Cas9 to a specific site in the genome that matches the target sequence, whereupon it cleaves the DNA, creating a double-stranded break (DSB). This can be repaired by non-homologous end joining, NHEJ (often introducing indels that can result in frameshift and gene knockout), or by homologous recombination, HR (which allows the integration of additional

sequences or basepair changes at that site if a homologous repair template / donor sequence is also introduced).[81–85]

This CRISPR/Cas9 system can be adapted for use in loss-of-function genetic screens with both positive and negative selection approaches. Instead of using individual sgRNAs against specific genes of interest, a genome-wide or targeted sgRNA library is used, introduced into cells as a lentiviral pool expressing thousands of sgRNAs (with multiple sgRNAs per target gene, plus a number of non-targeting control sgRNAs)[77–79]. The Cas9 protein can be expressed in the same constructs as the sgRNAs in the library[77]; alternatively, cell lines can be generated that stably express Cas9 in a conditional or constitutive manner[78,79], and these target cells can then be lentivirally transduced with the sgRNA pool. The lentiviral infection step is done at a low multiplicity of infection to ensure only single-copy integration of the sgRNA constructs into target cells (in contrast to the sgRNA overexpression involved in smaller scale gene editing). Once stably integrated into the genome, individual sgRNA sequences can be detected, and the number of cells carrying each sgRNA counted, by high-throughput sequencing.[77–79]

Therefore, "hit" genes can be identified at the end of a genetic screen by comparing the difference in sgRNA abundance between treated and untreated samples, or between final samples and initial day 0 samples. This reveals sgRNAs that are depleted or enriched (for negative or positive selection screens respectively) relative to non-targeting control sgRNAs in the library. This information can then be used to calculate a score for each gene in the library, highlighting important "hit" genes for which most of the sgRNAs have shown a change in abundance. In a negative selection screen, an sgRNA will be depleted if cleavage and subsequent knockout of its target gene confers a selective disadvantage in the majority of cells expressing that construct. Meanwhile in a positive selection screen, an sgRNA will be enriched if cleavage and knockout of its target gene confers a selective advantage.[77–79]

Transcription factors as therapeutic targets in ccRCC

The genetic intratumour heterogeneity in ccRCC[50,51] may contribute to the development of drug resistance, a recurring problem in ccRCC treatment. Targeting driver mutations that are ubiquitously present across ccRCC primary tumours and metastases may thus provide a therapeutic advantage[50,51].

Targeting VHL inactivation is an attractive therapeutic option – a method of restoring VHL therapeutically, for instance via some form of "VHL-mimetic", could be beneficial[25]. However it is difficult to directly target the inactivation of tumour suppressor genes[50,86]. Alternatively, recent work has demonstrated the efficacy of a small molecule inhibitor of the HIF2α transcription factor[75,76]; unfortunately, ccRCC cell lines exhibit variable sensitivity to this HIF2α antagonist, and can develop drug resistance after prolonged treatment[75,76].

In an attempt to identify therapeutic targets, various genetic screens have been carried out in VHL-deficient ccRCC cells, with a focus on identifying potential kinase, phosphatase, and metabolic targets[87–90]. However, I hypothesised that the phenotypes of renal cancer cells are fundamentally defined by core transcriptional networks, and our aim is to identify these key transcriptional programmes in ccRCC.

Cellular identity, behaviour, and responses to stimuli are all outcomes of carefully regulated gene expression programmes. This is exemplified by the key factors that define stem cell properties[91], and also in the context of cancer where studies have shown that several key developmental transcription factors determine glioblastoma cancer stem cell phenotypes[92–94].

Transcription factors act as effectors for the outputs of multiple signalling pathways, including those implicated in tumourigenesis and tumour progression[86,95]; thus, altered transcription factor function has the potential to drive cancer development. In the context of ccRCC transcriptional networks, deregulated HIF-driven transcription plays a critical role; but it is unknown as to whether this pathway single-handedly drives ccRCC development or whether it is just one part of a more complex network that defines ccRCC phenotypes. Indeed, the observation in the literature that ccRCC cell lines can become resistant to the recently developed HIF2α antagonist[75,76] could suggest that pathways other than VHL-HIF are involved.

Although transcription factors have generally been considered undruggable, there is evidence to support the potential efficacy of targeting transcription factors in cancer[86,95]. Notable examples include targeting nuclear steroid receptors in breast and prostate cancer, suppressing the expression of the transcription factor MITF in drug-resistant BRAF-mutant melanoma[96], and, in the context of ccRCC, the recently developed HIF2α antagonist that has shown efficacy *in vitro* and *in vivo*, despite previous conviction that this RCC-associated transcription factor was undruggable[75,76].

I have thus conducted an *in vitro* CRISPR/Cas9-based genetic depletion screen in VHL-/- ccRCC cell lines using a targeted single-guide RNA (sgRNA) library targeting 50 RCC-associated transcription factors. With this approach I hope to identify transcription factors that are essential in ccRCC, and to further characterise their roles in these tumours. Ultimately, this work will have implications for the potential value of these transcription factors as therapeutic targets.

HNF1B and HNF1B-associated disease

Hepatocyte nuclear factor 1-beta (HNF1B) is a homeodomain-containing transcription factor encoded by the *HNF1B* gene on chromosome 17q12[97–102]. It is a member of the hepatocyte nuclear factor-1 (HNF1) family, along with hepatocyte nuclear factor 1-alpha (HNF1A)[101,103,104]. HNF1B can act as a transcriptional activator or repressor, and binds to DNA as either a homodimer or a heterodimer with HNF1A; both recognise the same consensus sequence, 5'-GTTAATNATTAAC-3' [97,100–104]. HNF1B is expressed in a number of epithelial organs, particularly in the kidney but also in the liver, pancreas, urogenital tract, intestine, lung and thymus[97–101,103,104]. HNF1B regulates tissue-specific gene expression in the epithelia of these different organs, and plays a role in epithelial organ development and maintenance[97–101,103,104]. In the kidney, HNF1B is expressed in tubular epithelial cells in all segments of the nephron, including the proximal and distal tubules, loops of Henle, and renal collecting ducts[101,103].

Various mutations in *HNF1B* have been identified. The most common mutation is complete deletion of the *HNF1B* gene, which occurs due to a chromosomal microdeletion at 17q12 that results in the loss of *HNF1B* and 14 other genes[97,98]. *HNF1B* deletion accounts for roughly half of all cases of HNF1B-related disease[97,98]. The 17q12 region is prone to genomic rearrangements, and partial deletions of *HNF1B* have also been detected[97]. The other half of cases involve heterozygous mutations in the *HNF1B* gene, in the form of base substitutions and small insertions or deletions that occur in the coding region or splice sites[97]. These mutations tend to occur within the first four of *HNF1B*'s nine exons, with mutational hotspots in exon 2, exon 4, and the splice site in intron 2.[97] So far there has been no evidence for any genotype-phenotype correlation, suggesting that HNF1B-related disease arises due to haploinsufficiency[97,99]. When inherited, *HNF1B* mutations (both deletions and mutations in the

coding region or splice sites) follow an autosomal dominant pattern of inheritance; however, in up to 50% of cases these various *HNF1B* mutations arise spontaneously[97–99].

Based on observations from individuals and families with inherited *HNF1B* mutations, the phenotypes that develop can be highly variable between different patients, even within the same family or between families that carry the same mutation[97,99]. Additional genetic events, environmental influences, and possibly also random variation in *HNF1B* expression during development may thus play a role in determining the precise effects of *HNF1B* mutation[97,99].

In humans, heterozygous mutations in *HNF1B* result in a variety of renal and extra-renal phenotypes[97–100]. Perhaps due to the fact that HNF1B is expressed in multiple epithelial organs, the phenotypes produced by *HNF1B* mutations can manifest in a number of different organs including the kidney, pancreas, genital tract, liver and intestine[97–99]. Abnormalities can occur in a single organ in isolation, or can develop in multiple organs[97,98]. In general, the effects of *HNF1B* mutations first develop in the kidneys, and indeed congenital renal abnormalities are often attributed to *HNF1B* mutations[97,98].

HNF1B mutations can give rise to a type of early-onset diabetes mellitus known as maturity-onset diabetes of the young, type 5 (MODY5), which is "an autosomal dominant form of noninsulin diabetes mellitus"[98] caused by pancreatic β-cell dysfunction and subsequent impaired insulin secretion[97–101,103,104]. In addition to pancreatic endocrine dysfunction (manifesting as diabetes), HNF1B mutations can occasionally lead to pancreatic exocrine dysfunction and pancreatic malformations (such as pancreatic hypoplasia, atrophy, or agenesis)[97–99].

MODY5 is typically associated with renal malformations and/or abnormalities in other organs (such as those of the urogenital tract)[97–101,103,104]. The most commonly associated renal malformation is the development of kidney cysts, a combination referred to as renal cysts and diabetes (RCAD) syndrome[97–101,103,104]. A range of renal cystic abnormalities may occur, including "simple cysts, multicystic renal dysplasia, and hypoplastic glomerulocystic kidney disease"[101, 97–100,103,104].

HNF1B mutations in humans can also result in a number of other renal malformations, such as: renal agenesis, aplasia, or hypoplasia; oligomeganephronia; familial juvenile hyperuricaemic nephropathy; single, horseshoe, or duplex kidneys; hydronephrosis and

hydroureter, and tubular dysfunction (the latter suggesting that HNF1B also plays a role in the maintenance of functional tubules, in addition to renal development).[97–101,103,104]

The process of organ development is conserved between mammals and several other species, such as zebrafish and *Xenopus laevis*. Different animal models have therefore been used to study the effects of HNF1B deficiency on organ development.[97]

In mice, homozygous deletion of *Hnf1b* is embryonic lethal[98], while heterozygous deletion of *Hnf1b* does not produce a phenotype (in contrast to humans)[97,98,101,105,106]. Murine studies of HNF1B-related disease therefore require the use of transgenic mouse models that involve, for instance, conditional *Hnf1b* knockout in specific tissues or at specific stages of organ development[98]. Studies in such models have shown that Hnf1b plays an essential role in pancreas development, consistent with the pancreatic dysfunction and diabetes seen in humans with HNF1B mutations[97,98].

Transgenic mouse studies have also shown that deletion of *Hnf1b* leads to various defects in renal development, depending on the developmental time point or renal tissue type in which *Hnf1b* is deleted, indicating that HNF1B is also critical for several stages of renal development[97,98]. For example, inactivation of *Hnf1b* in the metanephric mesenchyme (which develops into the nephron) impairs the expansion and differentiation of the renal tubules, resulting in the formation of aberrant nephrons lacking the proximal and distal tubules and the loop of Henle. This suggests that HNF1B is important for tubule development in the nephron.[97–99]

Moreover, *Hnf1b* inactivation or deletion or the expression of a dominant-negative *Hnf1b* mutant, specifically in the mouse kidney, results in the formation of renal cysts, as well as increased cell proliferation and dilatation of the ureter and renal tubules[97–101,103,107,108]. This is consistent with the human data on the renal malformations that develop in patients with *HNF1B* mutations.

Similarly, zebrafish expressing different mutant alleles of *vhnf1* (the *HNF1B* homologue in zebrafish) exhibit the formation of renal cysts, as well as underdevelopment of the pancreas and liver[97]. Meanwhile, in *X. laevis*, different *HNF1B* mutations were found to impair the development of the pronephros (the embryonic kidney), to varying extents[97]. *Hnf1b* is also one of 4 transcription factors shown to be capable of reprogramming mouse and human fibroblasts into cells that resemble renal tubular epithelial cells[109].

HNF1B mutations in humans can also lead to the development of chromophobe renal cell carcinoma (ChRCC), a rare subtype of RCC that accounts for 4-5% of RCC tumours (in contrast to ccRCC, which accounts for 75-85% of RCC tumours)[1,2,97–100,104]. Studies have identified biallelic inactivation of *HNF1B* in several ChRCC cases with associated MODY5: patients exhibited germline mutation of one *HNF1B* allele, with somatic deletion of the other allele in the ChRCC tumours themselves, suggesting that *HNF1B* was inactivated via the classic Knudson two-hit mechanism in these ChRCC cases[97,98,100,104].

An immunohistochemical (IHC) study performed with samples of different renal tumour types also revealed a lack of HNF1B expression in the majority of ChRCC samples, suggesting that the absence of HNF1B might contribute to ChRCC tumourigenesis and that HNF1B may function as a tumour suppressor in this subtype of RCC. While *HNF1B* mutations were not detected in sporadic ChRCC tumours, these lesions still exhibited a lack of HNF1B protein expression. By contrast, HNF1B expression was still found in the majority of samples for other renal tumour types analysed in this IHC study, namely renal oncocytoma (which is benign), papillary renal cell carcinoma (PRCC), and ccRCC. The underexpression of HNF1B thus seems to be a particular feature of ChRCC tumours.[104]

Consistent with this, another study investigating *HNF1B* inactivation in ovarian carcinoma and three different RCC subtypes (renal oncocytoma, ChRCC and ccRCC) found that ChRCC tumours could exhibit bilalleic inactivation of *HNF1B* and also concluded that HNF1B might act as a tumour suppressor in ChRCC. Similarly, this study did not find evidence of HNF1B inactivation in ccRCC and ovarian tumours.[100]

Other studies have implicated HNF1B overexpression in ovarian clear cell carcinoma[97,99,102,110–112]. HNF1B genetic variants have also been proposed to influence the risk of prostate cancer, (albeit with contrasting results in the literature[113–115]), as well as the risk of endometrial cancer[116].[97,99,117–121]

Overall, there is substantial evidence in the literature demonstrating the importance of HNF1B in organ development, particularly that of the kidney. There is also ample human data and data from animal models implicating HNF1B mutations or deletions in a variety of renal and extra-renal diseases, which manifest in the various epithelial organs in which HNF1B is normally expressed. Furthermore, HNF1B has been implicated in a variety of different epithelial tumours, with different potential roles depending on tumour type: HNF1B may act as either a tumour

suppressor that is inactivated, or as a highly expressed driver of tumourigenesis. Thus, the role of HNF1B in cancer may be context-dependent, and is perhaps influenced by other genetic or epigenetic events specific to the tumour type in question. These observations in the context of cancer, along with the evidence demonstrating the variety of organs affected and phenotypes produced by HNF1B mutations, could suggest that HNF1B typically functions as part of lineage-specific transcriptional networks that influence its role in different tissues.

Approach & Rationale

ccRCCs have a tendency to develop drug resistance, rendering current treatment strategies (such as kinase inhibitors, anti-angiogenic drugs, and even the recently developed HIF2A inhibitor) ineffective in the long-term. Effective treatment of ccRCC thus requires new therapeutic targets, ideally targets that are ubiquitously essential across all tumour subclones and represent truncal tumour dependencies.

The first part of the overall hypothesis behind my PhD research is that transcription factors may represent ideal therapeutic targets, as they are critical regulators of cell survival and cell phenotypes, including in the context of cancer. They might thus be more likely to be essential for tumour cell survival across all tumour subclones. Moreover, transcription factors are often at the endpoints of signalling pathways, in some cases with multiple signalling pathways converging on the same transcription factors. Therapeutically targeting a transcription factor thus has the potential to inhibit multiple driver pathways in tumour cells simultaneously. This approach might also be less likely to be circumvented by the redundant functions of other pathway components or transcription factors, potentially hindering the development of drug resistance.

The second part of the overall hypothesis behind my PhD research is that ccRCC phenotypes may be fundamentally defined by tissue-specific core transcriptional networks. As detailed in the sub-sections above, there is evidence in the literature suggesting that lineage-specific transcription factors can play important roles in tumourigenesis and tumour progression. Similarly, ccRCCs may also depend on lineage transcription factors that are not necessarily mutated, but which may drive transcriptional programmes that are critical to support tumour growth and survival.

The two aims of my PhD are thus as follows:

- Aim 1: to identify additional and potentially novel transcription factor dependencies in ccRCC.
- Aim 2: to characterise the mechanisms by which these essential transcription factors support ccRCC.

To achieve Aim 1, I conducted a CRISPR/Cas9-based genetic depletion screen *in vitro*, with an sgRNA library targeting 50 transcription factor genes that are particularly highly expressed in RCC. This screen highlighted HNF1B as a candidate transcription factor that played an essential role in ccRCC.

To achieve Aim 2, I carried out *in vitro* proliferation assays and *in vivo* tumour growth assays to determine the phenotype associated with HNF1B depletion in ccRCC. I subsequently performed assays to characterise the nature of this phenotype, as well as RNA-seq and ChIP-seq experiments to identify transcriptional targets of HNF1B that might be involved in its critical function in ccRCC. Finally, I investigated the potential contributions of MYC and BCL2 (which turned out to be HNF1B target genes in ccRCC) to HNF1B's importance in ccRCC cells.

Ultimately, the overall aim of my PhD project was to try to reach a comprehensive understanding of the mechanisms by which HNF1B supports ccRCC.

RESULTS

CHAPTER 1: *In vitro* CRISPR-Cas9 genetic depletion screen for ccRCC cell dependencies

1.1) CRISPR-Cas9 genetic depletion screen workflow

To identify additional and potentially novel ccRCC cell dependencies, I conducted a CRISPR/Cas9-based genetic depletion screen *in vitro*[77–79].

For the purposes of the screen I used a conditional Cas9 expression system (specifically a doxycycline (dox)-inducible system), as this allowed more precise temporal control over gene knockout. Therefore, this enabled me to avoid early knockout while expanding the cells for the screen after sgRNA library transduction, and to then induce gene knockout only at the very start of the screen with the addition of dox.

With a constitutive system however, the Cas9 would be constantly active and thus gene knockout would already be underway prior to the start of the screen. The cells would therefore experience a negative selection pressure prematurely. This would risk selecting for: cells that have lost or silenced Cas9 or sgRNA expression, cells that have not formed knockout mutations (i.e. cells that did not form insertions/deletions (indels) at all, or only formed in-frame mutations that did not result in protein knockout), or resistant cells (i.e. cells that have compensated for the effects of gene knockout). If such cell populations are generated before or during the early stages of the screen, they would continue to survive and proliferate, and this could introduce false negatives in the screen results.

I thus used 786-M1A-TGL ccRCC cells that had previously been transduced in the lab with a dox-inducible Cas9 expression construct (pCW-Cas9), and sorted into single cell clones. I used a clone that had one of the highest levels of Cas9 expression upon dox treatment (***Fig 2***). The ccRCC cell line 786-M1A-TGL is a derivative of the 786-O primary ccRCC cell line used in other RCC studies, but with a higher metastatic propensity[4]. The cells also contain a TGL transgene, which comprises the cDNA sequences of herpes simplex virus type 1 thymidine kinase (HSV1-tk), eGFP and firefly luciferase[122]. Cells can thus be distinguished *in vitro* by eGFP expression, and can be detected *in vivo* upon administration of luciferin (tumour growth can thus be quantified via bioluminescence measurements in live mice).

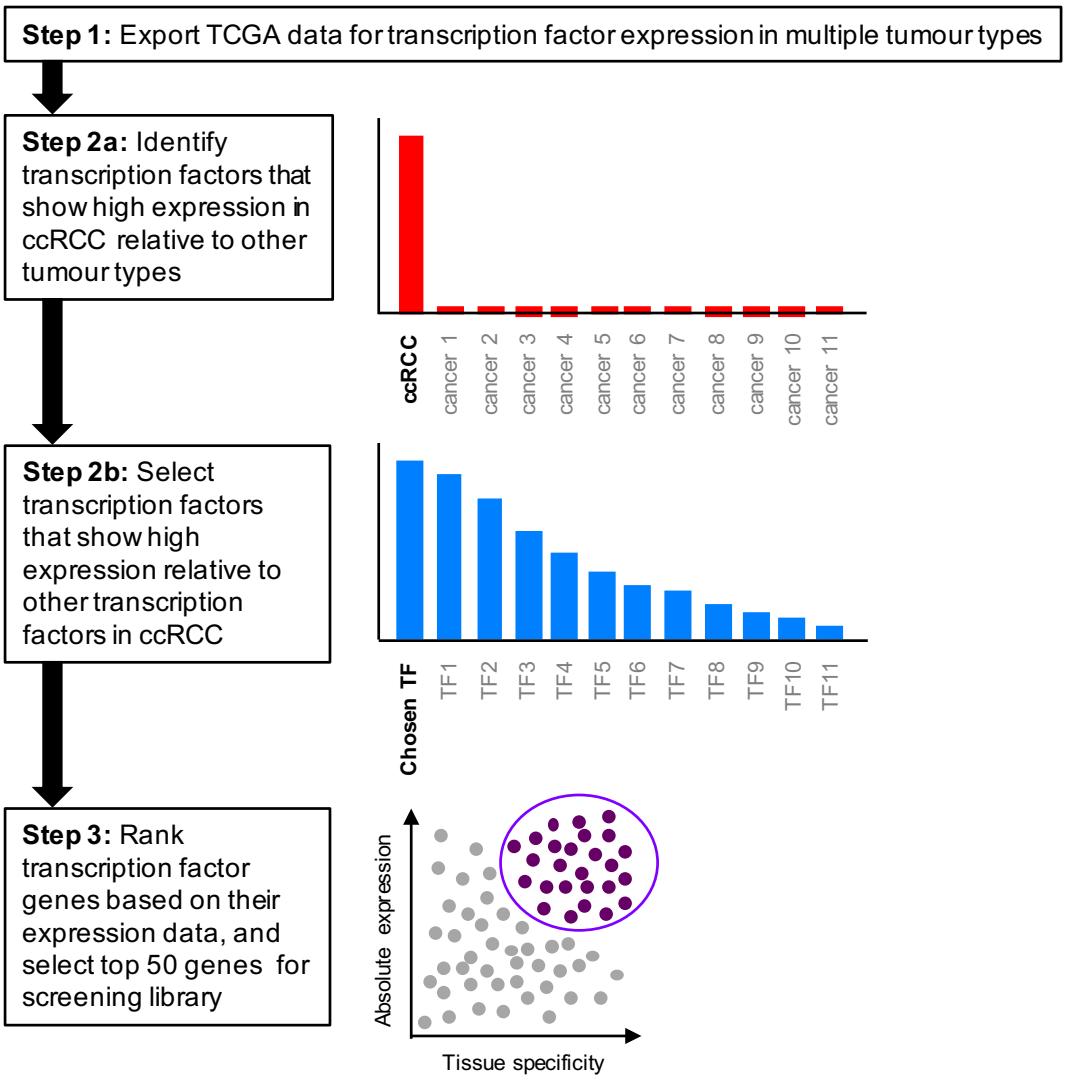


Fig 1: ***Screen library target gene selection.*** The 50 transcription factor target genes were chosen based on TCGA data for transcription factor mRNA expression in multiple tumour types. This dataset was first analysed to identify those transcription factors that were both highly expressed in ccRCC relative to other tumour types, and highly expressed within ccRCC relative to other transcription factors. These transcription factors were then ranked based on their expression levels, and the top 50 most highly expressed transcription factor genes were selected as targets for the screen sgRNA library.

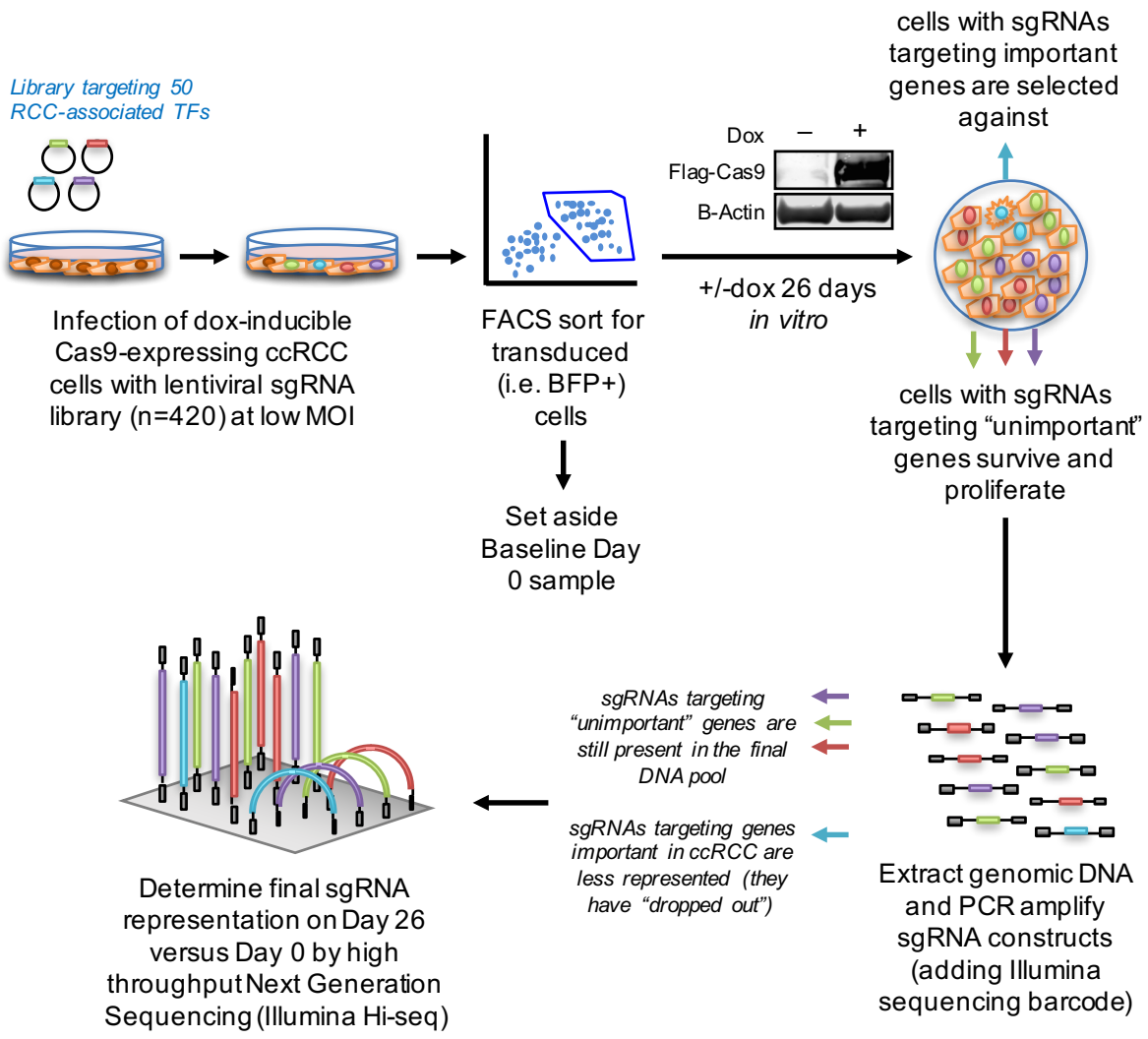


Fig 2: ***CRISPR/Cas9 genetic depletion screen workflow.*** Schematic illustrating the workflow for the CRISPR/Cas9 genetic depletion screen. I infected 786-M1A-TG-tet-on-Cas9 cells with the lentiviral sgRNA library targeting 50 transcription factor genes. The library contained a total of 420 sgRNAs, each in a BFP fluorescent backbone. I transduced the cells at a low multiplicity of infection (MOI) to ensure that each cell received only one sgRNA construct. I then sorted for BFP-positive (transduced) cells, retained a portion of these cells for my Baseline Day 0 sample, and divided the remainder into 2 samples: one +doxycycline (dox) and one –dox. I cultured the cells +/-dox for 26 days. At the end of the dox treatment period, I pelleted the cells and extracted genomic DNA from the Baseline Day 0, Day 26 +dox, and Day 26-dox cell samples. I then PCR amplified the sgRNA constructs from each genomic DNA pool, adding Illumina sequencing barcodes in the process. The final sgRNA representation in each sample was then assessed via high throughput Next Generation Sequencing (Illumina Hi-seq). I compared the sgRNA abundance between the Day 26 +dox and Day 0 samples to determine which sgRNAs had been depleted at the end of the screen. sgRNAs targeting "important" genes (and the cells expressing them) would be relatively less abundant in the final population at the end of the screen compared to Day 0 (such sgRNAs would appear to have "dropped out").

I transduced the clonal 786-M1A-TGL-tet-on-Cas9 cells with a lentiviral library comprising 420 CRISPR single guide RNAs (sgRNAs), targeting 50 RCC-associated transcription factors (8 sgRNAs per target, with 20 non-targeting control sgRNAs) (***Fig 1***, ***Fig 2***). The sgRNA sequences were taken from the whole-genome sgRNA library designed by Wang *et al*.[78].

The 50 transcription factor target genes were identified using TCGA data for transcription factor mRNA expression in multiple tumour types[53,54]. This dataset was first analysed to identify the subset of transcription factors that were both highly expressed in ccRCC relative to other tumour types, and highly expressed within ccRCC relative to other transcription factors. These transcription factors were then ranked based on their expression levels, and the top 50 most highly expressed transcription factor genes were selected as targets for the screen sgRNA library (***Fig 1***).

I transduced the 786-M1A-TGL-tet-on-Cas9 cells at a low multiplicity of infection (MOI) so that the majority of cells only received a single sgRNA (***Fig 2***, ***Fig 3B***), thus avoiding multiple gene knockouts (***Fig 2***). Each sgRNA construct within the library also contained a BFP fluorescent marker[79], allowing me to sort for successfully transduced cells based on BFP expression (***Fig 2***, ***Fig 3A***). 26% of the population was BFP-positive after transduction; I isolated these cells by FACs sorting, and retained and expanded them for the screen (***Fig 3A***).

The MOI can be calculated based on a Poisson distribution for the probability of a target cell to absorb a given number of virus particles when infected at a given MOI. This has been adapted in the graph in ***Figure 3B***, which plots the fraction of cells with a single virus integration against the total percentage of positive cells obtained. Thus, for 26% BFP-positive cells, roughly 86% of the transduced cells received only a single sgRNA construct (***Fig 3B***).

I divided the sorted BFP-positive 786-M1A-TGL-tet-on-Cas9 cells into a baseline day 0 sample, one dox-treated sample, and one untreated control sample (***Fig 2***). I froze down multiple cell pellets from the Baseline Day 0 cells for later analysis. I then cultured the dox-treated and untreated samples *in vitro* for 26 days, with and without 1ug/ml dox respectively.

During this 26-day culture period, dox treatment would induce Cas9 expression, facilitating gene knockout in the 786-M1A-TGL-tet-on-Cas9 cells (***Fig 2***). Cells with sgRNAs targeting important genes in ccRCC would be selected against, whereas cells with sgRNAs targeting “unimportant” genes would survive and proliferate. Thus, within the dox-treated

sample, there would be a reduced proportion of cells that had been expressing sgRNAs targeting important genes, and a higher proportion of cells that had been expressing sgRNAs targeting "unimportant" genes. As a result, there would be a decreased representation of sgRNAs targeting important genes within the final cell population (compared to both the original sgRNA representation in the Day 0 population, and the sgRNA representation in the untreated Day 26 – dox sample).

I expanded the cells for each sample over the 26-day treatment period, ensuring that the total number of transduced cells per sample in the screen was at least 1000 times the total number of constructs (i.e. 420 sgRNAs) in the screen library, in order to avoid loss of library representation over the course of the screen. This minimum ratio of cell number to library size was based on a paper by Zuber *et al*.[123], in which the authors conducted RNAi screens using more than 1000 cells per shRNA, and maintained this minimum cell number throughout their experiment to avoid losing library representation. Indeed, Zuber *et al.* found an approximately 3 Log difference between the most and least abundant shRNAs in their library[123]. The dox medium was refreshed and the cells were passaged every 3 days, at a low splitting ratio of either 1:3 (for the dox-treated sample) or 1:4 (for the untreated sample).

FACs analysis at the end of the 26-day period confirmed that the BFP-positive (i.e. sgRNA-expressing) population remained stable in both the dox-treated and untreated samples, with over 90% of the cells expressing BFP (***Fig 4A-B***).

At the endpoint of the screen, I froze down multiple cell pellets from the Day 26 dox-treated and untreated samples for analysis. I then extracted genomic DNA from duplicate pellets for each of the three screen samples (Day 26 dox-treated ("Day 26 +dox"), Day 26 untreated ("Day 26 –dox"), and Baseline day 0). Each of these genomic DNA pools would contain the sgRNA constructs for that sample, which would then have to be isolated for sequencing. I thus PCR amplified the sgRNA constructs in each of the six genomic DNA samples (***Fig 2***). This PCR step served to add sample-specific barcodes to allow the different samples to be distinguished, as well as Illumina sequencing adapters required for next generation sequencing (NGS, specifically Illumina Hi-seq) of the sgRNA construct amplicons (***Fig 2***).

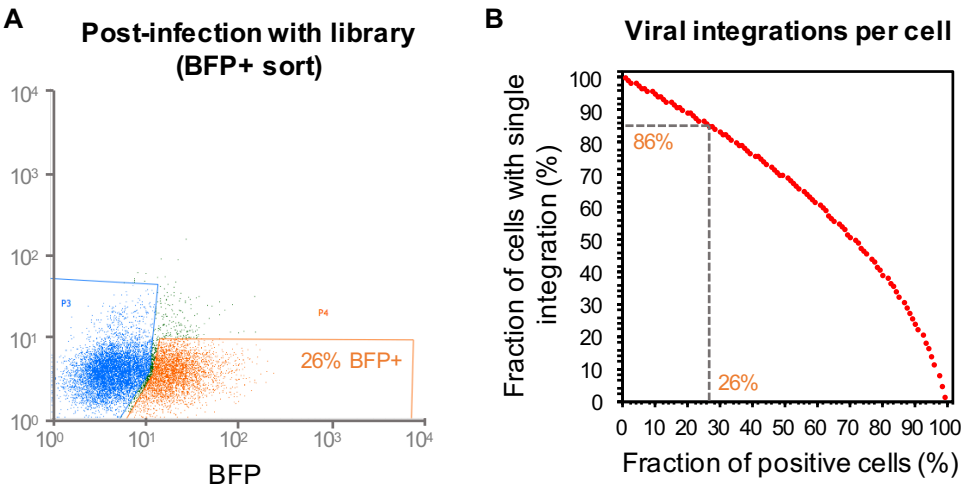


Fig 3: ***The 50 transcription factor sgRNA library was lentivirally transduced at a low MOI.*** (A) FACs plot showing the BFP+ fraction of lentivirally transduced cells after infection with the CRISPR/Cas9 sgRNA library. This 26% BFP+ population was isolated by cell sorting. A portion was retained as the Day 0 baseline sample, and the remainder was cultured +/-dox for 26 days for the screen, to induce Cas9 expression and thus gene knockout.

(B) Graph showing the relationship between the fraction of positive (i.e. successfully transduced) cells and the fraction of cells with a single integration in the population during a viral infection. This relationship gives an approximation of the MOI and thus the number of viral integrations per cell in a lentiviral transduction. Thus, having 26% positive cells (in this case BFP-positive) meant that roughly 86% of the transduced cells had a single integration, and thus contained only one sgRNA construct.

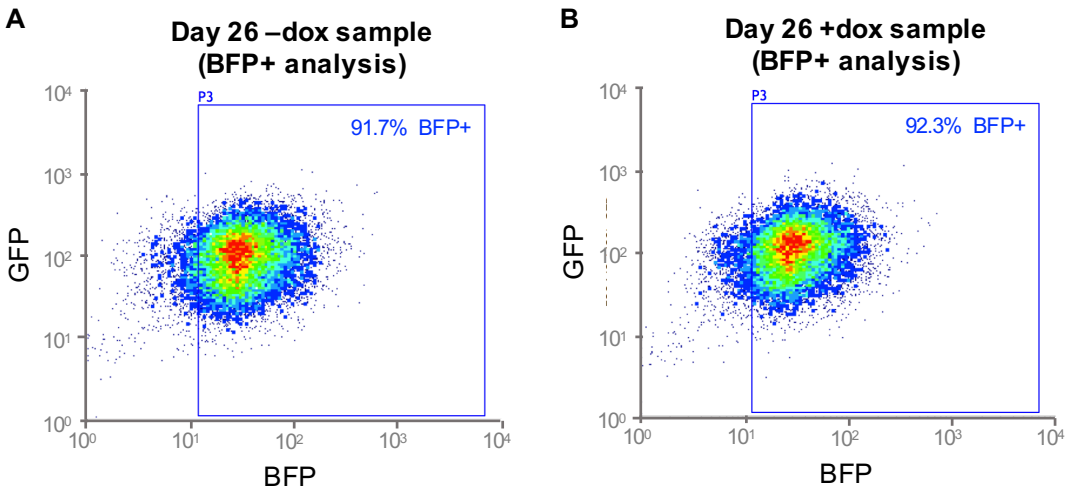


Fig 4: ***Library sgRNA expression was not lost by the end of the screen.*** FACs plots showing the percentage of BFP+ cells at the end of the screen in (A) the Day 26 –dox sample and (B) the Day 26 +dox sample, after being cultured for 26 days either without dox or with 1ug/ml dox respectively.

1.2) Preparing screen samples for Illumina sequencing

The original protocol and conditions for these screen PCRs had been developed previously in the lab. I helped to further optimise this original screen PCR protocol from a two-step PCR to a single-step PCR, in order to minimise the chances of PCR contamination.

The single-step PCRs were carried out with forward and reverse primers that comprised: (1) regions complementary to the vector sequences flanking each sgRNA in the library constructs (a.k.a the "constant regions"); (2) the sample-specific barcode sequence to differentiate between each of the Day 26 +dox, Day 26 –dox, and D0 baseline duplicate samples; and (3) the P5 or P7 (for forward and reverse primers respectively) Illumina sequencing adapters required for Illumina Hi-seq.

Because each sample had to be tagged with a unique barcode, each of my six DNA samples was amplified with a different pair of primers (*see Materials and Methods*). To generate sufficient quantities of each PCR product, multiple PCR reactions were performed for each sample (and thus with each primer pair). The replicate PCR reactions for each sample were then pooled at the end of the PCR step, prior to phenol-purification and SPRI-bead size exclusion. Only one water (H_2O) control reaction was performed for each primer pair used.

To prepare the PCR amplified DNA samples for Illumina Hi-seq, I first checked the PCR products and PCR H_2O controls for any contamination by running a small (2ul) aliquot of each pooled PCR, as well as the entire H_2O-only PCR reactions, on agarose gels (***Fig 5A***). Note that the gels in ***Figure 5A*** show both the PCR reactions for my duplicate samples (6 in total), as well as two samples belonging to another postdoc in the lab who was also performing screens at that time.

The DNA agarose gels confirmed that there was no PCR contamination in the H_2O control reactions (the first four lanes in each gel), and that each of the single-step PCR reactions (the last four lanes in each gel) generated a specific product of the correct size (274bp), in sufficient quantities for further purification (***Fig 5A; arrows***).

The pooled PCR products were then purified by phenol-purification, and size-selected by SPRI-bead size exclusion. We originally attempted to size-select the PCR products with gel extraction; however, we consistently lost too much yield. I thus helped to optimise the protocol to incorporate SPRI-bead size-exclusion instead.

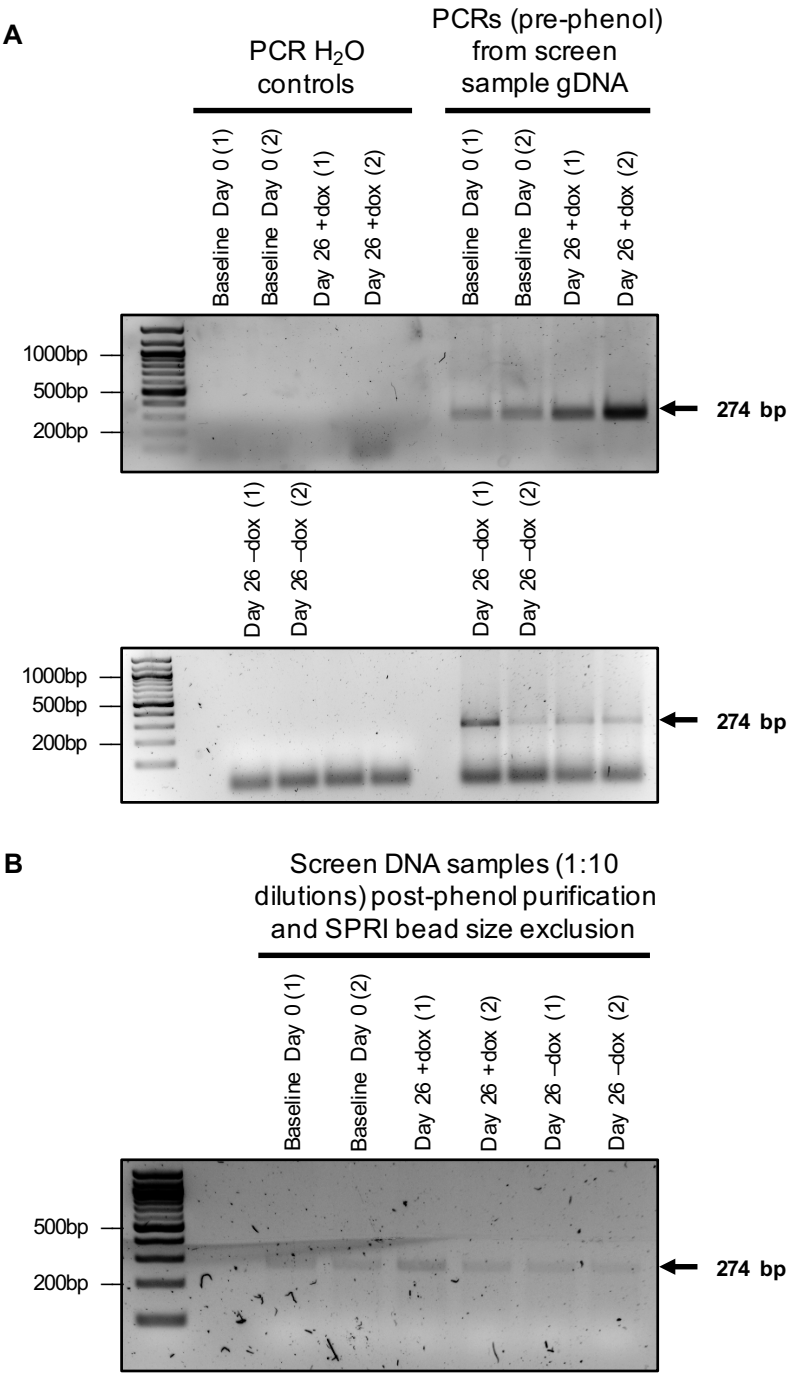


Fig 5: ***Quality control checks for screen sample PCR products before and after purification and size-selection.***

(A) DNA agarose gels showing the PCR H_2O controls and the PCR products generated from each screen sample (Baseline Day 0, Day 26 +dox and Day 26 –dox, in duplicate), prior to phenol-purification and SPRI-bead size exclusion. No PCR contamination can be seen in the H_2O control lanes, and the PCR step generated a specific product of the correct size (274bp).

(B) DNA agarose gel showing the final PCR products for each duplicate sample, after phenol purification and SPRI-bead size exclusion. There is still sufficient product remaining, and the size exclusion step successfully isolated the product of the correct size (274bp).

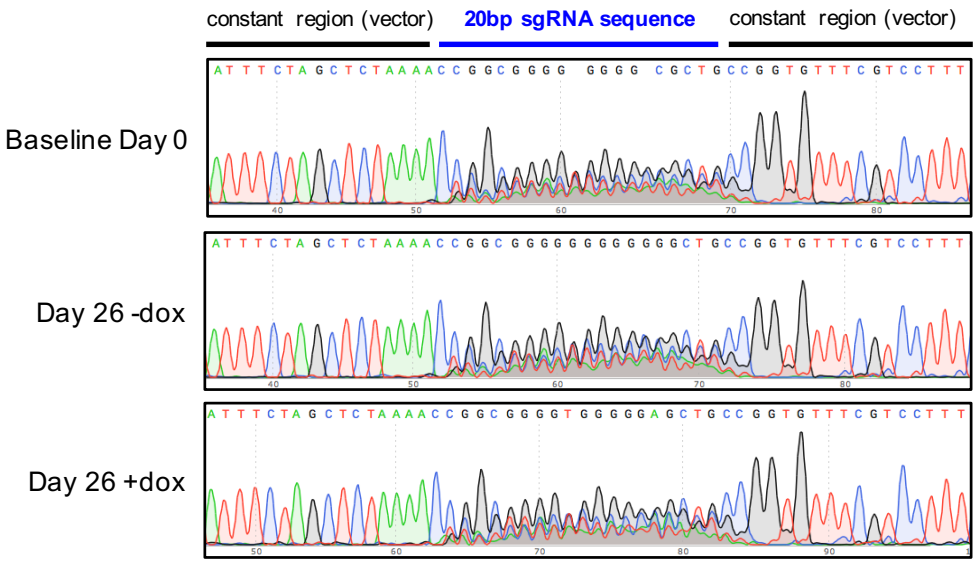


Fig 6: ***Sanger sequencing of the final PCR products for each screen sample, after amplification, purification and size-selection.***
Sanger sequencing data for the final DNA samples obtained after PCR amplifying and purifying the sgRNA constructs from the genomic DNA pools of each screen sample. Representative sequencing data is shown for only one duplicate per screen sample. The Sanger sequencing traces showed no sign of contamination, and the sequencing results showed constant regions upstream and downstream that matched the vector backbone sequences that flank each sgRNA, with a variable 20bp region in between that corresponded to the position of the sgRNA sequences in each library construct.

I carried out the phenol-purification and SPRI-bead size-exclusion for each of my 6 screen samples. I then made 1:10 dilutions of each of my samples, and ran an aliquot of each diluted sample on an agarose gel (***Fig 5B***) to confirm that the purification and size-selection steps had successfully isolated product of the correct size, and in sufficient quantities (i.e. without losing too much DNA yield) for subsequent sequencing library preparation and NGS.

The DNA agarose gel showed that all 6 of my samples contained PCR product of the correct size (274bp), with no additional or contaminating products of different sizes (***Fig 5B; arrow; lanes 2-7***). Moreover, the fact that bands were clearly visible for each sample, despite running an aliquot of a 1:10 dilution, indicated that I had retained sufficient quantities of DNA product after purification and size-selection for sequencing library preparation (***Fig 5B; lanes 2-7***).

Finally, I quantified my final screen samples by Qubit quantification, with the Qubit dsDNA HS (High Sensitivity) Assay Kit (*Invitrogen / Life Technologies*). I then sent aliquots of my final screen samples (post-PCR amplification, purification, and size-selection) for Sanger sequencing (***Fig 6***) to confirm that I had successfully isolated the sgRNA constructs from the genomic DNA pools for each screen sample.

The Sanger sequencing results for each sample showed sequences upstream and downstream that matched the sequences of the vector backbone that flank the sgRNAs in each library construct. These vector sequences were identical between the Baseline Day 0, Day 26 – dox and Day 26 +dox DNA samples, and were the only non-variable, constant regions in the sequence traces across all the duplicates (as expected, since the vector backbone is the same for all the sgRNA constructs in the library).

The sequencing results for each screen sample also showed a 20bp region that was highly variable in sequence, lying in between the constant regions. This corresponded to the position of the 20bp sgRNA sequences in each library construct, and thus indicated the presence of multiple 20bp sgRNA sequences in each screen DNA sample. Indeed, it is this variable 20bp region that was subsequently de-convoluted via Illumina sequencing to reveal individual sgRNAs.

Importantly, the Sanger sequencing traces also showed no sign of contamination in the final screen samples. Note that ***Figure 6*** shows only one representative sequence trace per screen sample, but that this sequencing data is representative of both duplicates for each of the Baseline Day 0, Day 26 –dox and Day 26 +dox DNA screen samples.

Therefore, the Sanger sequencing data confirmed that my final screen samples did indeed contain sgRNA constructs. Having verified this, the sequencing libraries were subsequently prepared with my screen samples by another member of the lab, and sent for Illumina Hi-seq to determine the sgRNA representation in each sample.

1.3) Analysis and interpretation of the Illumina sequencing data: a potential role for HNF1B

The Illumina sequencing data was analysed in-house as described in the *Materials and Methods*. To align and quantify the sgRNA sequences in each of the 6 screen samples, a custom bowtie index was created using the sequences of all 420 sgRNA constructs in the screen library. The raw Illumina sequencing data was then aligned to this index of library sgRNA sequences using bowtie1[124], and the number of reads per sgRNA were counted. This gave the raw counts for each sgRNA, which were then normalised to the total number of reads per sample. These normalised counts were then taken as a measure of sgRNA abundance.

The normalised counts per sgRNA were calculated in this way for each duplicate sample separately. I then compared these normalised counts between duplicates of the same screen sample to determine the consistency of the Illumina sequencing data between sample replicates. In other words, I compared duplicate 1 versus duplicate 2 of the Baseline Day 0 sample (***Fig 7A***), duplicate 1 versus duplicate 2 of the Day 26 +dox sample (***Fig 7B***), and duplicate 1 versus duplicate 2 of the Day 26 –dox sample (***Fig 7C***).

These comparisons are illustrated in the scatter plots in ***Figure 7***. In each scatter plot, each data point represents one of the 420 sgRNAs in the screen, with non-targeting controls highlighted in black, sgRNAs for the transcription factor HNF1B highlighted in red, and all other sgRNAs in grey. The axes represent the normalised counts per sgRNA (as an indicator of sgRNA abundance) for the duplicate sample indicated. The sgRNA normalised counts were highly consistent between duplicate samples, with most if not all of the data points lying along the XY diagonal between the axes (***Fig 7A-C***). The R coefficients were also close to 1 for each comparison: R = 0.9872 for the Baseline Day 0 duplicates (***Fig 7A***), R = 0.9867 for the Day 26 +dox duplicates (***Fig 7B***), and R = 0.9881 for the Day 26 –dox duplicates (***Fig 7C***). Thus, there was minimal variation in sgRNA abundance between the duplicates of each screen sample, indicating that the screen data was robust and reproducible.

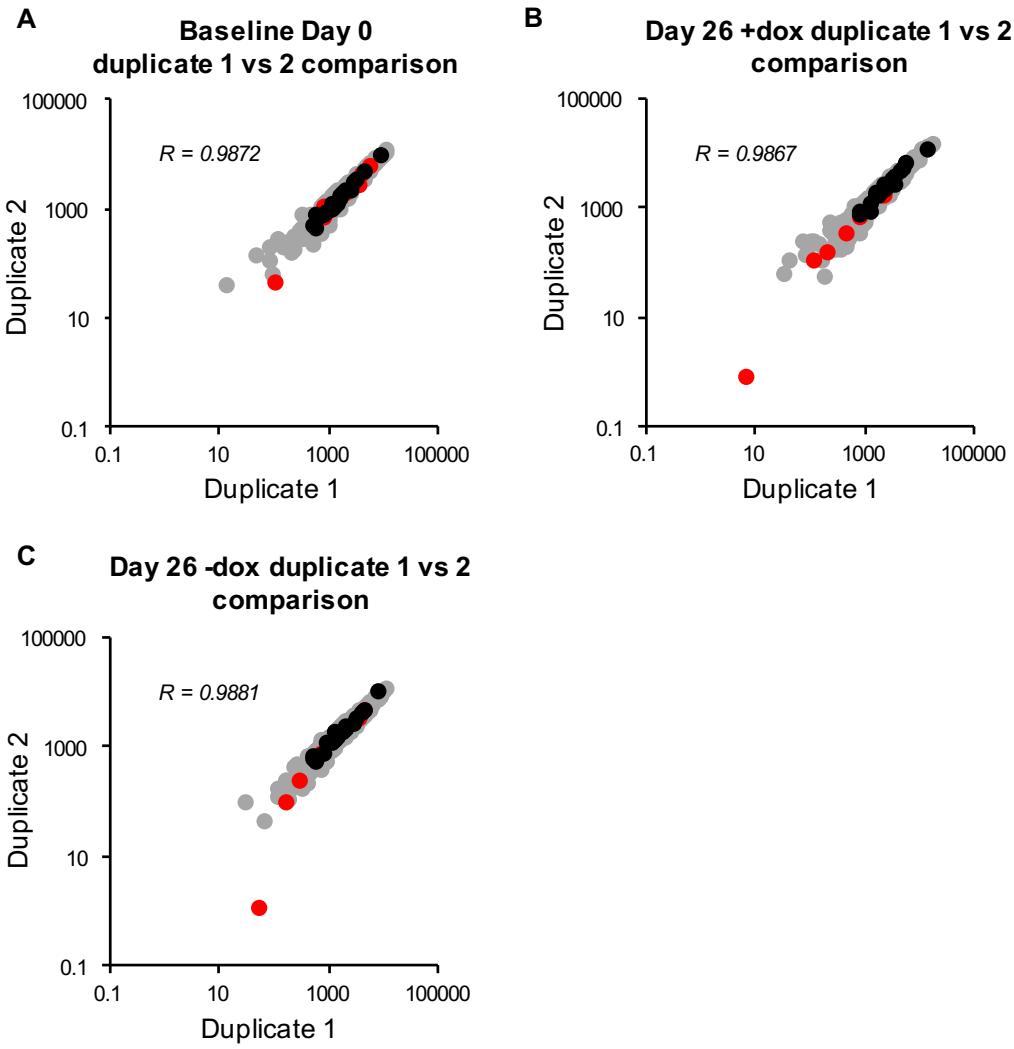


Fig 7: ***Comparisons of the sgRNA abundance between sample duplicates to determine the consistency of the Illumina sequencing data.*** Scatter plots showing the comparisons of the sgRNA abundance between the duplicates for (A) Baseline Day 0, (B) Day 26 +dox and (C) Day 26 –dox. In each dot plot, the data points represent the 420 individual sgRNAs in the screen, with non-targeting controls in black, HNF1B sgRNAs in red, and all other sgRNAs in grey. The axes represent the normalised counts per sgRNA (as an indicator of sgRNA abundance) for the duplicate sample indicated.

After comparing the sgRNA distributions between the screen sample duplicates, I compared the sgRNA distributions between the screen samples themselves. Comparing the sgRNA representations between the Day 26 +dox and the Baseline Day 0 samples identified sgRNAs that were under-represented (i.e. which had been depleted, or "dropped out") in the final sgRNA pool of the Day 26 +dox sample, relative to the original sgRNA pool in the Baseline Day 0 sample.

I also compared the final sgRNA representations between the Day 26 +dox and Day 26 –dox samples, and the final sgRNA representation in the Day 26 –dox sample against the initial representation in the Baseline Day 0 sample.

Comparing the final sgRNA representations between the Day 26 +dox and Day 26 –dox samples would also highlight sgRNAs that had been depleted in the treated versus untreated samples. The sgRNA constructs depleted, and the extent to which these constructs dropped out at the end of the screen, should be similar between the Day 26 +dox versus Day 26 –dox comparison and the Day 26 +dox versus Baseline D0 comparison, provided that there was no leakiness of Cas9 expression, and thus no background sgRNA loss, in the untreated cells that were not exposed to dox.

Meanwhile, comparing the sgRNA distributions between the Day 26 –dox and the Baseline Day 0 samples would reveal more definitively whether the dox-inducible Cas9 transgene was leaky or not. If there was little to no leakiness in transgene expression, then there would be minimal to no loss of sgRNA representation seen in the Day 26 –dox sample relative to the Baseline Day 0 sample. However, if there was a background level of leaky Cas9 expression without dox treatment, then one would expect to observe some sgRNA depletion in the Day 26 –dox sample relative to the Baseline Day 0 sample.

To facilitate these comparisons, I first averaged the normalised counts between the duplicates of each sample for each of the 420 sgRNAs, and compared these average normalised counts between (i) Day 26 +dox and Baseline Day 0 (***Fig 8A***), (ii) Day 26 +dox and Day 26 –dox (***Fig 8B***), and (iii) Day 26 –dox and Baseline Day 0 (***Fig 8C***).

In each scatter plot in ***Figure 8A-C***, the data points represent the 420 individual sgRNAs in the screen, with non-targeting controls highlighted in black, HNF1B sgRNAs highlighted in red, and all other sgRNAs in grey. The axes represent the average normalised counts for each sgRNA, for the indicated screen sample.

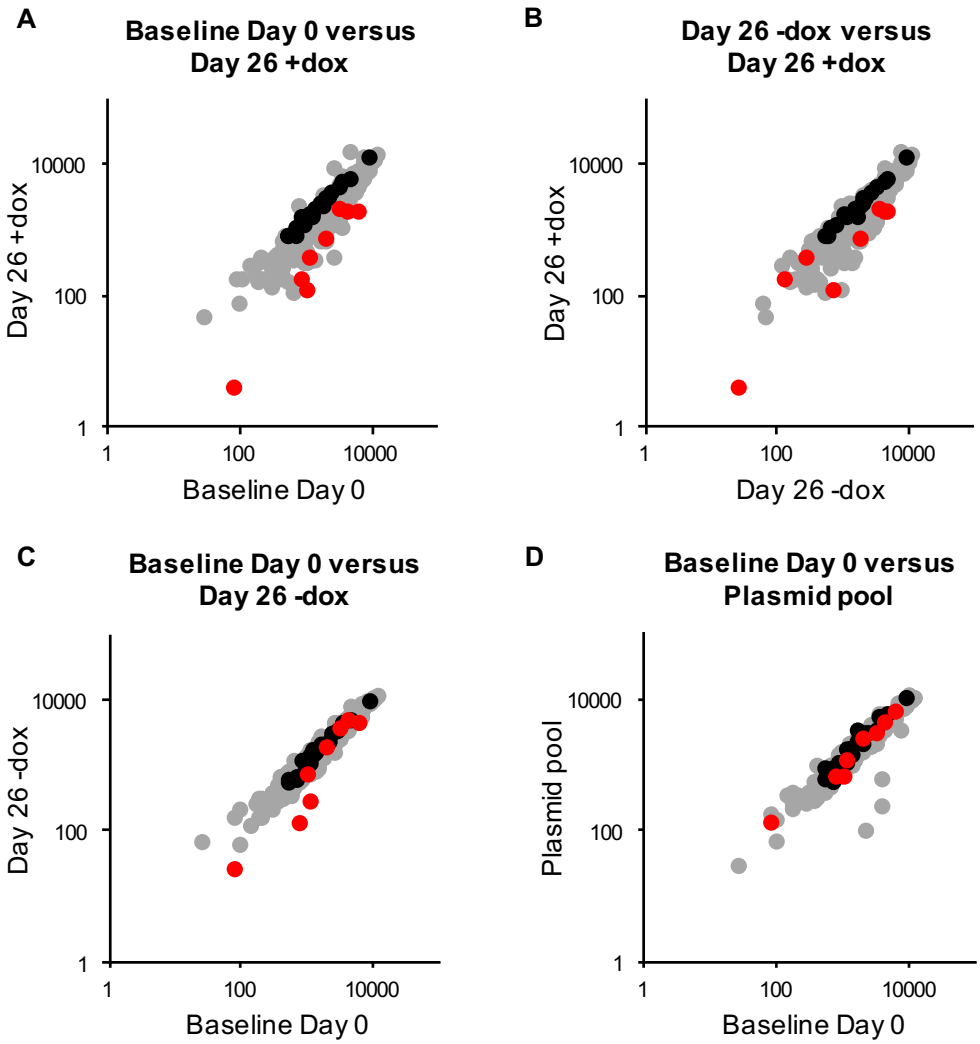


Fig 8: ***Comparisons of the sgRNA representation of all 420 sgRNAs between the original Plasmid pool, Baseline Day 0, Day 26 +dox, and Day 26 –dox screen samples.*** Scatter plots comparing the sgRNA representation in (A) Baseline Day 0 versus Day 26 +dox, (B) Day 26 –dox versus Day 26 +dox, (C) Baseline Day 0 versus Day 26 –dox, and (D) Baseline Day 0 versus the original plasmid pool. In each scatter plot, the data points represent the 420 individual sgRNAs in the screen, with non-targeting controls highlighted in black, HNF1B sgRNAs highlighted in red, and all other sgRNAs in grey. The axes represent the sgRNA abundance, given as average normalised counts (the raw counts per sgRNA normalised to the number of reads and averaged between duplicates) for the screen sample indicated, or given as normalised counts in the case of the plasmid pool in (*D*) (which was only sequenced as a single replicate).

In all three plots, the non-targeting control sgRNAs (black) seem to lie along the XY diagonal, suggesting that they were not substantially depleted or enriched between samples over the course of the screen. Thus, they are unlikely to have off-target effects.

Meanwhile, there is a clear downwards shift in multiple targeting sgRNAs (grey and red) when comparing Day 26 +dox with Baseline Day 0 (***Fig 8A***), and also when comparing Day 26 +dox with Day 26 –dox (***Fig 8B***). This downwards shift indicates that these sgRNAs are less represented with dox treatment (and thus with Cas9 expression and gene knockout), and suggests that these constructs most likely target transcription factor genes that are important in ccRCC. Some of the most substantially depleted sgRNAs are those targeting the renal developmental transcription factor HNF1B (highlighted in red).

Comparing the sgRNA representation in the Day 26 –dox sample against that of the Baseline Day 0 sample revealed minimal sgRNA dropout in the absence of dox: fewer sgRNAs were depleted, and to a much lesser extent, compared to the frequency and extent of sgRNA depletion observed in Day 26 +dox versus Baseline Day 0 (***Fig 8C***).

Additionally, there is still a clear difference between the sgRNA representations in the Day 26 +dox sample versus the Day 26 –dox sample (***Fig 8B***), emphasizing that the loss of sgRNA representation in the Day 26 +dox sample exceeded that in the Day 26 –dox sample.

Both comparisons thus suggest that the Cas9 expression in the 786-M1A-TGL-tet-on-Cas9 cells was only slightly leaky. Therefore, minimal gene knockout would have occurred prior to the start of dox treatment (i.e. prior to the start of the screen), and there would have been limited loss of sgRNA representation resulting from background Cas9 expression to obscure the biological effects of target gene knockout.

Finally, I compared the sgRNA distribution of the Baseline Day 0 screen sample (i.e. the sgRNA representation at the start of the screen) against the original sgRNA distribution of the lentiviral plasmid library itself (i.e. the sgRNA representation of the library plasmid pool *before* the start of the screen) (***Fig 8D***).

The lentiviral sgRNA library was sequenced post-synthesis (as a single replicate), prior to producing virus and infecting the 786-M1A-TGL-tet-on-Cas9 cells with the 420 library constructs. This sequencing data (and thus the normalised counts calculated) for each of the 420 sgRNAs in the library plasmid pool therefore provided a reference point for the earliest, original distribution of sgRNA constructs, prior to cell transduction and thus *before* the start of the

screen. In other words, the sgRNA distribution in the plasmid pool indicated the most complete representation of each sgRNA in the newly synthesised library, before any sgRNA depletion could have occurred. Meanwhile, the sequencing data from the Baseline Day 0 screen sample (and thus the average normalised counts calculated across the sample duplicates) represented the sgRNA distribution at the start of the screen, after virus generation and cell transduction with the lentiviral library. Comparing the sgRNA distributions between the Baseline Day 0 screen sample and the original plasmid pool would thus reveal whether there was any loss of sgRNA representation between the point of library synthesis and the generation of the Baseline Day 0 sgRNA pool.

The scatter plot in ***Figure 8D*** compares the average normalised counts of the Baseline Day 0 sample against the normalised counts of the original library plasmid pool, for each of the 420 sgRNAs. As with ***Figs 8A-C***, the data points in ***Fig 8D*** represent the 420 individual sgRNAs in the screen, with non-targeting controls highlighted in black, HNF1B sgRNAs highlighted in red, and all other sgRNAs in grey. The *x*-axis represents the average normalised counts per sgRNA for the Baseline Day 0 sample, while the *y*-axis represents the normalised counts per sgRNA for the original library plasmid pool.

The majority of the sgRNAs seem to lie along the XY diagonal, with only a few constructs appearing to be enriched in the Baseline Day 0 sgRNA pool relative to the library plasmid sgRNA pool. This suggests that there was minimal loss of sgRNA representation from that of the original plasmid pool post-synthesis, in the process of lentivirus production and cell transduction. Therefore, the screen started with an sgRNA representation that corresponded well to the original representation of each sgRNA construct in the lentiviral plasmid library.

I next calculated the fold changes in the average normalised counts per sgRNA for each sample comparison, by normalising: (i) the Baseline Day 0 averages to the plasmid pool normalised counts (as no replicate samples were available to take an average in the case of the latter) (***Fig 9A***), (ii) the Day 26 +dox averages to the Baseline Day 0 averages (***Fig 9B***), (iii) the Day 26 +dox averages to the Day 26 –dox averages (***Fig 9C***), and (iv) the Day 26 –dox averages to the Baseline Day 0 averages (***Fig 9D***). These calculations were once again performed for each of the 420 sgRNAs individually. Calculating the fold changes in this way gave a "depletion score" for each sgRNA construct. This allowed me to better determine the extent to which sgRNAs dropped out, and thus the relative differences in sgRNA representation between the start

of the screen (Baseline Day 0) and the library plasmid pool (***Fig 9A***), and also between the screen samples (Baseline Day 0, Day 26 +dox and Day 26 –dox; ***Fig 9B-D***) in each comparison.

Each waterfall plot in ***Figure 9A-D*** shows the extent of sgRNA depletion (represented as $\log_2$ fold change in average normalised counts on the *y* axis) of all 420 sgRNAs in the screen, for the indicated sample comparisons. The individual sgRNAs are arranged along the *x* axis in order of the most depleted to the least depleted (i.e. the most relatively enriched), with each bar representing a different sgRNA. Non-targeting controls are highlighted in black and sgRNAs for the transcription factor HNF1B are highlighted in red, with all other sgRNAs shown in grey.

Most, if not all, of the 20 non-targeting control sgRNAs (highlighted in black) appeared relatively enriched in each of the three waterfall plots comparing the screen samples (***Fig 9B-D***), but to varying extents. In the waterfall plot comparing Baseline Day 0 with the original plasmid pool, meanwhile, the non-targeting controls were more evenly distributed in terms of relative depletion or enrichment, with some appearing slightly enriched and the remainder appearing slightly depleted in the Baseline Day 0 sample relative to the plasmid pool (***Fig 9A***).

Indeed, when comparing the extent of sgRNA depletion between Baseline Day 0 and the plasmid pool, there was minimal sgRNA depletion or enrichment at the start of the screen relative to the original library post-synthesis (***Fig 9A***). This further indicates that there were minimal changes in sgRNA representation after the library was transduced into cells for the screen, thus suggesting that the screen began with an sgRNA representation similar to that of the original sgRNA plasmid library.

When comparing the extent of sgRNA depletion between the Day 26 +dox and Baseline Day 0 samples, sgRNAs targeting the transcription factor HNF1B (highlighted in red) were among the most substantially depleted sgRNAs in the screen, with all 8 HNF1B sgRNAs appearing within the 50 most depleted sgRNAs out of the 420 total constructs (***Fig 9B***).

Similarly, 6 out of 8 HNF1B sgRNAs were also within the 50 most depleted sgRNAs when comparing Day 26 +dox against Day 26 –dox (***Fig 9C***). Indeed, the extent of sgRNA dropout seen in the Day 26 +dox sample relative to the Day 26 –dox sample was similar to the extent of sgRNA dropout seen in the previous comparison of the Day 26 +dox sample versus the Baseline Day 0 sample (***Fig 9B-C***). A similar proportion of sgRNAs were depleted in the Day 26 +dox sample relative to both Day 26 –dox and Baseline Day 0. Moreover, the depletion scores for these under-represented sgRNAs in the Day 26 +dox sample were also similar relative to

both Day 26 –dox and Baseline Day 0. However, the maximum extent of depletion observed was slightly less when comparing Day 26 +dox against Day 26 –dox; the most depleted sgRNAs dropped out with a much lower depletion score when comparing Day 26 +dox against Baseline Day 0 (***Fig 9B-C***).

Comparing the extent of sgRNA depletion between the Day 26 –dox sample and the Baseline Day 0 sample revealed a low level of sgRNA depletion in the absence of dox (***Fig 9D***). However (as observed previously in ***Fig 8C***), fewer sgRNAs were depleted, and to a much lesser extent, compared to the frequency and extent of sgRNA depletion observed in Day 26 +dox versus Baseline Day 0. This suggests that there was some leakiness in expression of the dox-inducible Cas9 transgene that would trigger gene knockout in the untreated cells (***Fig 9B-D***).

Despite this, the loss of sgRNA representation in the Day 26 +dox sample still exceeded that in the Day 26 –dox sample (***Fig 9C***). Furthermore, the extent of sgRNA depletion seen in the Day 26 –dox sample relative to the Baseline Day 0 sample (***Fig 9D***) was considerably less than that seen when comparing both Day 26 +dox versus Baseline Day 0 (***Fig 9B***) and Day 26 +dox versus Day 26 –dox (***Fig 9C***). Therefore, although this slight leakiness would have resulted in a background level of Cas9 expression and gene knockout in the untreated cells, the low level of sgRNA dropout resulting from this background Cas9 activity was insufficient to detract from the sgRNA dropout resulting from the biological effects of these sgRNAs.

Thus, the observation that HNF1B sgRNAs were among the most substantially depleted sgRNAs in the screen is still valid, and is most likely due to perturbations in cell fitness caused by the HNF1B knockout mediated by these HNF1B sgRNAs.

The graph in ***Fig 9E*** is an excerpt of the data from ***Fig 9B*** for the extent of sgRNA depletion in Day 26 +dox versus Baseline Day 0, showing the depletion scores for the 8 HNF1B sgRNAs in isolation, along with the (relatively enriched) non-targeting control sgRNA #18 (CTRL0018, a.k.a NTC18) for comparison (***Fig 9E***).

In addition to examining the depletion scores per sgRNA, the median depletion score per gene (i.e. across all 8 sgRNAs per transcription factor gene) was also calculated, and the genes were then ranked from most to least depleted using a rank sum test (***Fig 9F***). This revealed transcription factors whose sgRNAs had dropped out significantly overall, relative to the non-targeting controls and other sgRNA constructs in the screen library, which would further suggest that those transcription factors might be important in ccRCC.

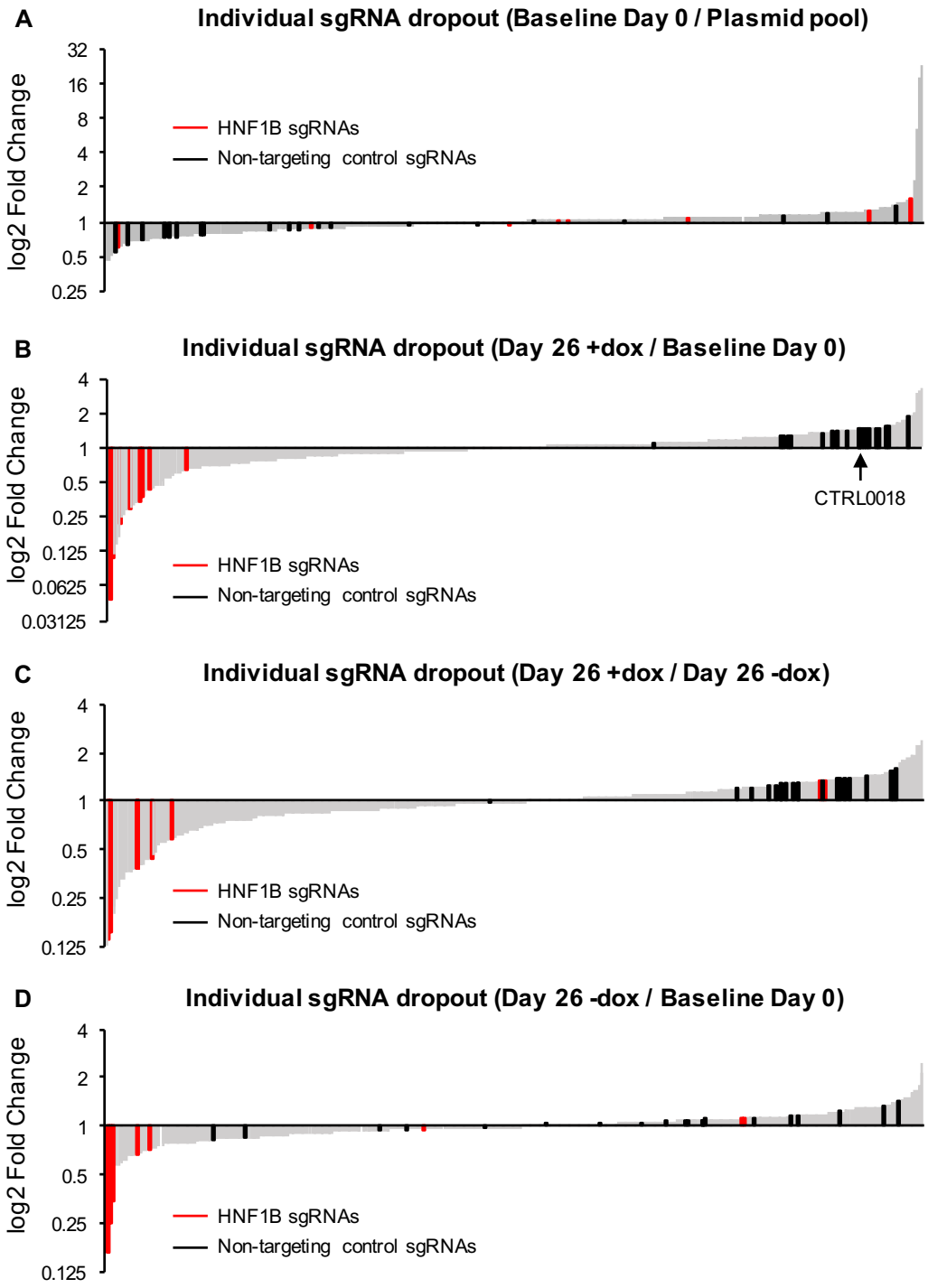


Fig 9: ***sgRNAs targeting HNF1B were among the most prominently depleted at the end of the screen.*** P.T.O for figure legend.

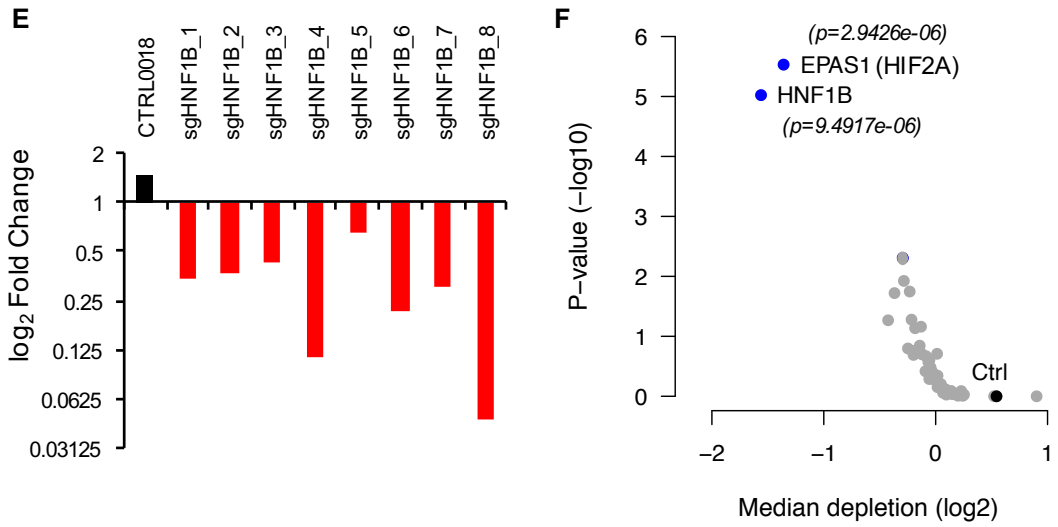


Fig 9: ***sgRNAs targeting HNF1B were among the most prominently depleted at the end of the screen.***

(A) Waterfall plot showing the $\log_2$ fold changes between the average normalised counts per sgRNA in the Baseline Day 0 screen sample and the normalised counts per sgRNA in the original plasmid pool, for all 420 sgRNAs in the screen. The individual sgRNAs are arranged along the *x* axis in order of the most depleted to the least depleted (i.e. the most relatively enriched), with each bar representing a different sgRNA. Non-targeting controls are highlighted in black and sgRNAs for the transcription factor HNF1B are highlighted in red, with all other sgRNAs shown in grey.

(B)-(D) Waterfall plots showing the extent of sgRNA depletion of all 420 sgRNAs in the screen, represented as $\log_2$ fold change in average normalised counts in (*B*) Day 26 +dox versus Baseline Day 0, (*C*) Day 26 +dox versus Day 26 –dox, and (*D*) Day 26 –dox versus Baseline Day 0. The individual sgRNAs are arranged along the *x* axis in order of the most depleted to the least depleted (i.e. the most relatively enriched), with each bar representing a different sgRNA. Non-targeting controls are highlighted in black and sgRNAs for the transcription factor HNF1B are highlighted in red, with all other sgRNAs shown in grey.

(E) Bar graph showing the extent of depletion (represented as $\log_2$ fold change in average normalised counts) of all 8 sgRNAs targeting HNF1B, along with the non-targeting control sgRNA #18 (CTRL0018) for comparison.

(F) Scatter plot showing the median depletion score per gene (represented as $\log_2$ median depletion), plotted against p-value (calculated by rank sum test, and represented as $-\log_{10}$ of the p-value). Each data point represents one of the 50 genes targeted in the screen library, with non-targeting controls highlighted in black and transcription factor genes in grey. The two highest ranking (i.e. most depleted) transcription factors, HIF2A (EPAS1) and HNF1B, are highlighted in blue, with their respective p-values indicated (EPAS1 p=2.9426e-06; HNF1B p=9.4917e-06).

The scatter plot in ***Figure 9F*** shows the median depletion score per gene (represented as $\log_2$ median depletion) along the x axis, and the p-value from the rank sum test (represented as $-\log_{10}$ of the p-value) along the y axis. Each data point represents one of the 50 genes targeted in the screen library, with the non-targeting controls highlighted in black and the transcription factor genes in grey.

Transcription factor genes with lower median depletion scores (i.e. with sgRNAs that were more strongly depleted overall) and small p-values (i.e. whose sgRNAs dropped out with greater statistical significance) would appear closer to the upper left of the graph. Meanwhile, those genes with higher median depletion scores (i.e. whose sgRNAs were relatively enriched or less strongly depleted), and larger p-values (i.e. whose sgRNAs dropped out with less statistical significance), would appear closer to the bottom right of the graph.

Once again, the non-targeting controls (black) appeared to be relatively enriched, with a high median depletion score. The transcription factor HIF2A (EPAS1) was one of the two most significantly depleted transcription factors out of the 50 transcription factor genes (*EPAS1* $p=2.9426 \times 10^{-6}$). As the importance of HIF2A in ccRCC tumourigenesis has already been established[55,64,71], this result was expected and acted as a positive control for the screen (***Fig 9F***). The other most significantly depleted transcription factor was HNF1B (*HNF1B* $p=9.4917 \times 10^{-6}$) (***Fig 9F***), which is consistent with the observation that all 8 HNF1B sgRNAs were strongly depleted in the screen in the Day 26 +dox sample relative to Baseline Day 0.

Taking this data together, the renal developmental transcription factor HNF1B was one of the strongest hits from the screen. The 8 sgRNAs targeting HNF1B were among the most depleted sgRNAs when comparing the representation of individual sgRNAs in the Day 26 +dox sample versus the Day 26 –dox and Baseline Day 0 samples. HNF1B also ranked as one of the top two most significantly depleted transcription factor genes in the screen based on its median depletion score. The fact that HNF1B dropped out in the screen so significantly suggests that this transcription factor plays an important role in supporting ccRCC. I thus selected HNF1B as the best hit for further characterisation and validation.

I first determined the HNF1B sgRNA knockout efficiency to identify the best constructs to use for further experiments. I transduced fresh 786-M1A-TGL-tet-on-Cas9 cells independently with either NTC18 or one of the 8 HNF1B sgRNAs from the screen. I then treated the transduced cells with 1ug/ml dox for 8 days to induce Cas9 expression and gene knockout,

and extracted whole cell protein lysates from each cell line to blot for HNF1B and check the knockout efficiency of each HNF1B sgRNA.

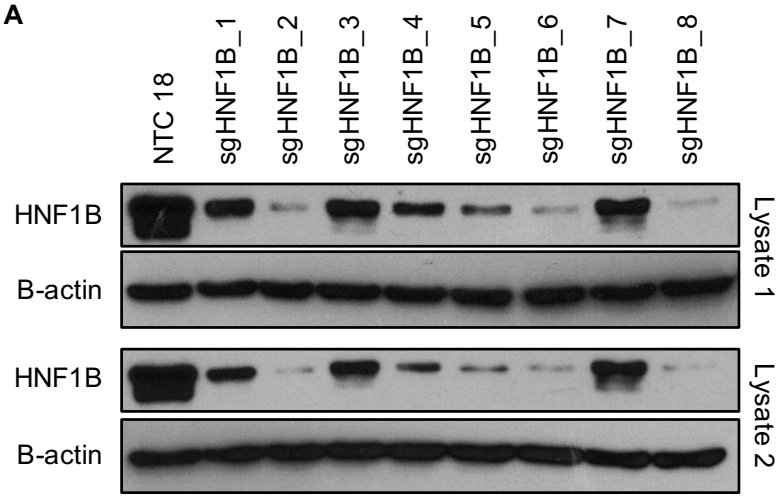
Western blots for HNF1B (from 2 independent lysates) showed that most of the 8 sgRNAs knockdown HNF1B with high efficiency, with half of the constructs mediating almost complete HNF1B knockout relative to the NTC18 control (***Fig 10A***). However, the extent of protein knockout mediated by each sgRNA (***Fig 10A***) did not necessarily correspond to the extent to which that sgRNA dropped out in the screen (***Fig 9E, 10B***).

For instance, HNF1B sgRNA #8 dropped out the most in the screen and also mediated the most efficient protein knockout. HNF1B sgRNAs #6 and #3, which were the third and seventh most depleted out of the eight constructs respectively, also gave the third and seventh highest efficiencies of protein knockout respectively. The other constructs, however, showed depletion scores that did not correspond to their knockout efficiency. For example, HNF1B sgRNA #2 ranked sixth out of the eight constructs in terms of depletion score, but ranked second in terms of protein knockout efficiency. Similarly, HNF1B sgRNA #5 ranked last in terms of depletion score in the screen, but still managed to confer an intermediate level of HNF1B protein knockout, ranking fourth out of the eight constructs for knockout efficiency. Meanwhile, HNF1B sgRNA #4 was the second most depleted in the screen but only ranked fifth in terms of protein knockout efficiency. (***Fig 9E, 10A-B***)

Based on a combination of protein knockout efficiency (***Fig 10A***) and sgRNA depletion score in the screen sequencing data (***Fig 9E***), I chose HNF1B sgRNAs #2 and #8 for further experiments and validation.

Both HNF1B sgRNAs #2 and #8 were highly efficient at knocking out HNF1B protein (***Fig 10A***). However, sgRNA #8 was strongly depleted in the screen and might thus incur additional off-target effects, whereas sgRNA #2 was much less strongly depleted in the screen and is thus less likely to have off-target effects (***Fig 9E***). Therefore, selecting these two HNF1B sgRNA constructs allowed me to validate HNF1B as a hit from the screen with two independent and effective sgRNAs, and also to compare the results seen with both constructs to determine whether any off-target effects could be observed in my validation experiments. If both sgRNAs produced consistent phenotypes, then this would strengthen the conclusions from my data. However, if the effects seen with sg8 differed and/or were stronger, this might suggest an off-target effect associated with the sg8 construct.

A

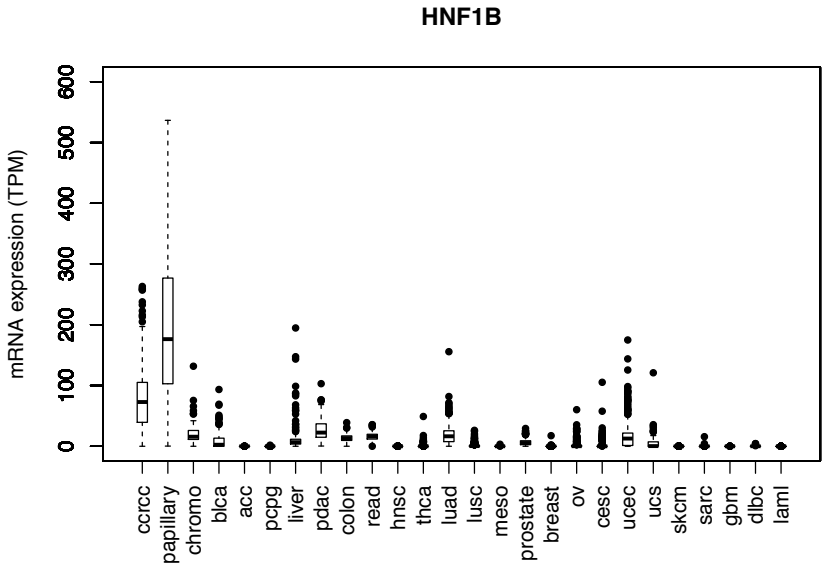


B

	Ranking in terms of screen depletion score	Ranking in terms of HNF1B protein knockout efficiency
1	sgHNF1B #8	sgHNF1B #8
2	sgHNF1B #4	sgHNF1B #2
3	sgHNF1B #6	sgHNF1B #6
4	sgHNF1B #7	sgHNF1B #5
5	sgHNF1B #1	sgHNF1B #4
6	sgHNF1B #2	sgHNF1B #1
7	sgHNF1B #3	sgHNF1B #3
8	sgHNF1B #5	sgHNF1B #7

Fig 10: ***HNF1B sgRNA knockdown / knockout efficiency.***
(A) HNF1B Western blots from 2 independent whole cell protein lysates, extracted from 786-M1A-TGL-tet-on-Cas9 cells transduced independently with NTC18 or one of the 8 HNF1B sgRNAs from the screen, and then treated with 1ug/ml dox for 8 days to induce Cas9 expression and gene knockout. Most of the 8 sgRNAs knockdown HNF1B with high efficiency, with half of the constructs mediating almost complete HNF1B knockout relative to the NTC18 control.

(B) Table showing the rankings of each of the 8 HNF1B sgRNAs from most to least depleted in the screen (*second column*) and from highest to lowest protein knockout efficiency (*third column*). Comparing these rankings revealed that the extent of protein knockout mediated by each sgRNA did not necessarily correspond to the extent to which that sgRNA dropped out in the screen (***see also Fig 10A versus Fig 9D***).



Code	Tumour type
ccrcc	Clear cell renal cell carcinoma
papillary	Papillary renal cell carcinoma
chromo	Chromophobe renal cell carcinoma
blca	Bladder urothelial carcinoma
acc	Adrenocortical carcinoma
pcpg	Pheochromocytoma and Paraganglioma
liver	Liver hepatocellular carcinoma
pdac	Pancreatic ductal adenocarcinoma
colon	Colon adenocarcinoma
read	Rectum adenocarcinoma
hnsc	Head and Neck squamous cell carcinoma
thca	Thyroid carcinoma
luad	Lung adenocarcinoma

Code	Tumour type
lusc	Lung squamous cell carcinoma
meso	Mesothelioma
prostate	Prostate adenocarcinoma
breast	Breast invasive carcinoma
ov	Ovarian serous cystadenocarcinoma
cesc	Cervical squamous cell carcinoma and endocervical adenocarcinoma
ucec	Uterine corpus endometrial carcinoma
ucs	Uterine carcinosarcoma
skcm	Skin cutaneous melanoma
sarc	Sarcoma
gbm	Glioblastoma multiforme
dlbc	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
laml	Acute Myeloid Leukemia

Fig 11: ***TCGA data for HNF1B mRNA expression in different tumour types.*** Dot plot showing TCGA data for HNF1B mRNA expression in multiple tumour types, including ccRCC (*top*). HNF1B mRNA expression levels are shown on the *y*-axis. The different tumour types are indicated along the *x*-axis. The TCGA tumour codes are explained in the table (*bottom*). Based on TCGA mRNA expression data, HNF1B is highly expressed at the mRNA level in ccRCC compared to other tumour types.

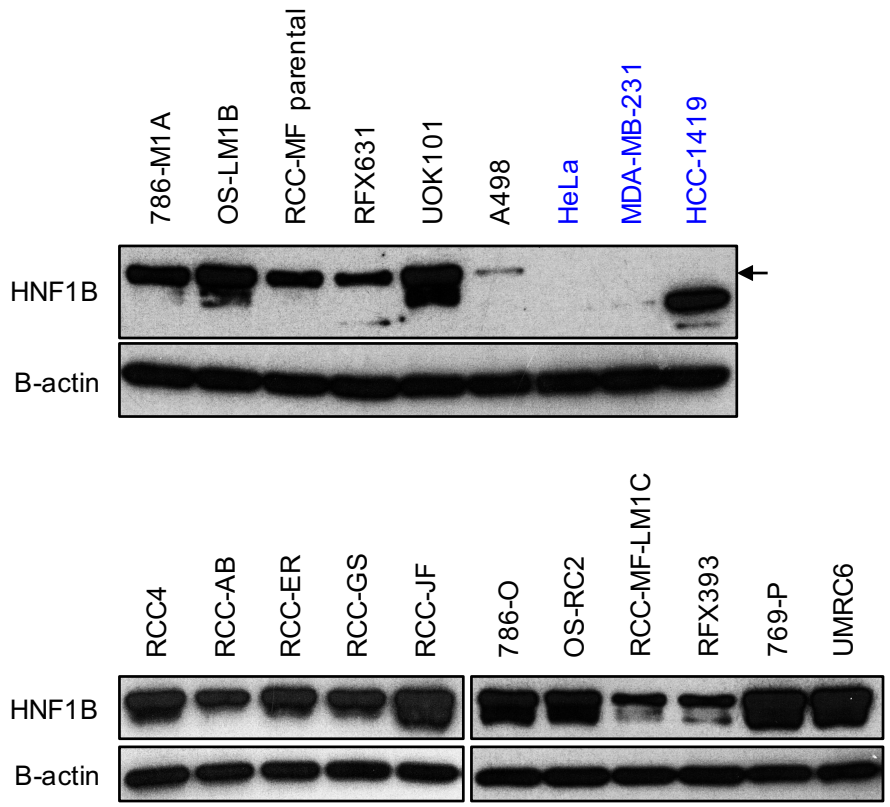


Fig 12: ***HNF1B protein expression in ccRCC cell lines versus other tumour cell types.*** HNF1B Western blots with whole cell protein lysates from 17 ccRCC cell lines (including 3 pairs of parental lines and their more metastatic derivatives: 786-O & 786-M1A, OS-RC2 & OS-LM1B, and RCC-MF & RCC-MF-LM1C, respectively), as well as HeLas and 2 breast cancer cell lines (MDA-MB-231 and HCC-1419). HNF1B was expressed in all 17 ccRCC cell lines, but was absent in HeLas and breast cancer cells.

CHAPTER 2: Establishing a phenotype for HNF1B knockout in ccRCC cells *in vitro* and *in vivo*

2.1) HNF1B is highly expressed in ccRCC tumours and cell lines

The next step was to confirm the phenotype for HNF1B knockout and thus validate the screen data placing HNF1B as one of the top "hits" out of the 50 transcription factors targeted, as well as verify whether HNF1B does indeed play an important role in ccRCC.

Based on TCGA mRNA expression data, HNF1B is particularly highly expressed at the mRNA level in ccRCC relative to other tumour types (***Fig 11***). Indeed, this is in keeping with how we selected the 50 transcription factor target genes for the sgRNA library. The high HNF1B expression levels in ccRCC may suggest that this transcription factor has a relatively tissue-specific role in these tumours.

To further verify the specificity and relative levels of HNF1B expression in ccRCC versus other tumour types, I blotted for HNF1B with whole cell protein lysates from 17 ccRCC cell lines (including 3 pairs of parental lines and their more metastatic derivatives: 786-O & 786-M1A, OS-RC2 & OS-LM1B, and RCC-MF & RCC-MF-LM1C, respectively)[4], versus 3 non-ccRCC cell lines (HeLas and 2 breast cancer cell lines, MDA-MB-231 and HCC-1419) (***Fig 12***). Consistent with the TCGA mRNA expression data, HNF1B protein is expressed at high levels specifically in ccRCC cell lines compared to non-ccRCC cancer cell lines (at the very least, based on the cell line panel used here). This further underscores the possibility that high HNF1B expression is specific to ccRCC (or at the very least found in only a limited range of tumour tissues including ccRCC), and that this may reflect a relatively tissue-specific importance for HNF1B in these tumours.

2.2) Investigating the phenotype for HNF1B CRISPR/Cas9 knockout in vitro

In light of the slight leakiness of the 786-M1A-TGL-tet-on-Cas9 clone used in the screen, and due to previous data in the lab (data not shown) demonstrating that clonal cell lines can exhibit impaired growth and aberrant behaviour *in vivo* that is inconsistent with that of parental cell lines or pooled cell populations, I decided to use a constitutive Cas9 expression system instead of the dox-inducible one to investigate the role of HNF1B in ccRCC.

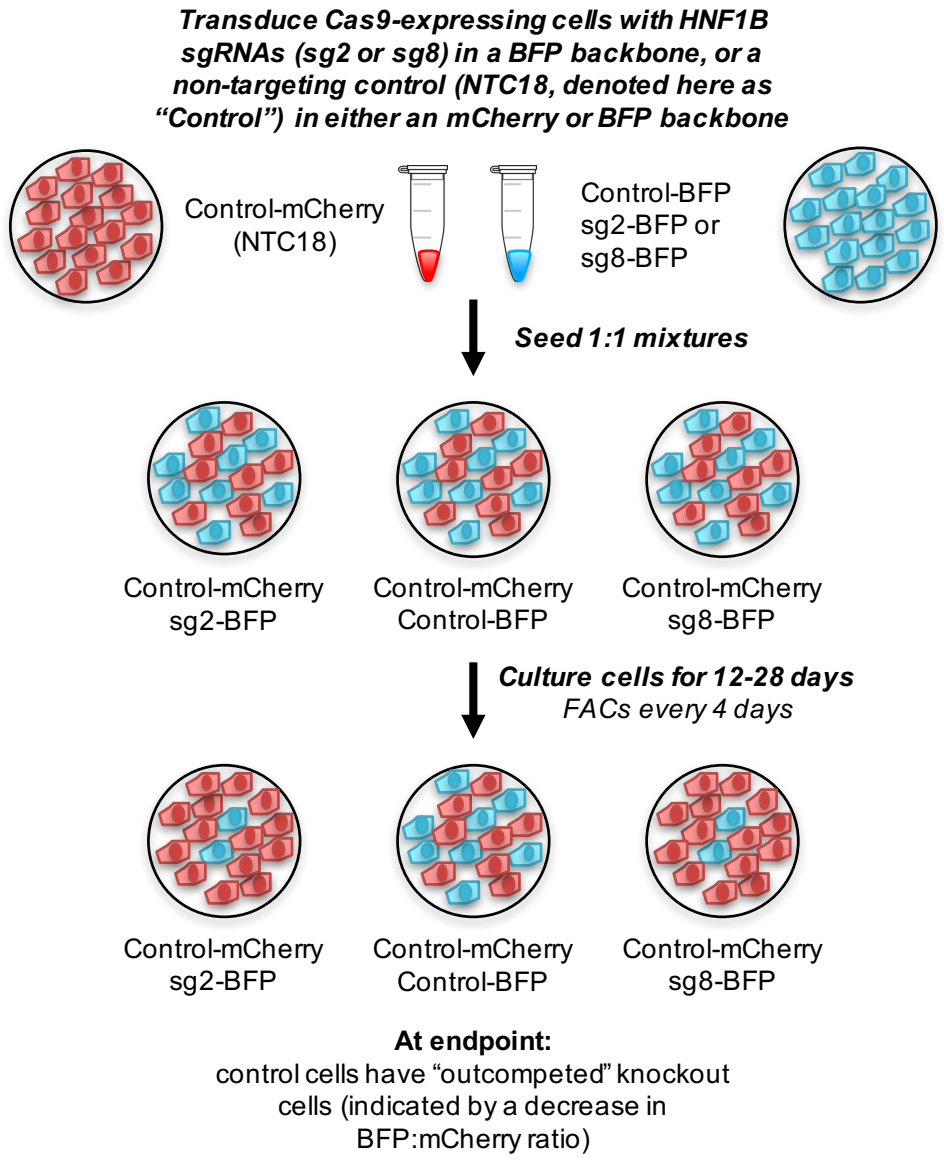


Fig 13: ***Competition proliferation assay workflow.*** Schematic illustrating the workflow of the competition proliferation assays used to validate the phenotype of HNF1B knockout *in vitro*. ccRCC cells expressing a constitutive Cas9 expression construct (lenti-Cas9-blast) were transduced with either NTC18 (a.k.a Control), HNF1B sgRNA #2 (sg2) or HNF1B sgRNA #8 (sg8), in either an mCherry or BFP fluorescent backbone.
Control-mCherry cells were mixed in a 50:50 ratio with either Control-BFP (as the control group), sg2-BFP or sg8-BFP cells. These cell mixtures were cultured for several weeks. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the BFP:mCherry ratio over time. If HNF1B knockout resulted in a selective disadvantage, the sgRNA-expressing cells would either die or proliferate less than the controls, and would be outcompeted by the control cells in the assay mixture. Thus, by the endpoint of the assay, the percentage of sg2-BFP or sg8-BFP cells would be greatly reduced relative to the percentage of Control-mCherry cells (and the BFP:mCherry ratio would have decreased). By contrast, the proportion of BFP versus mCherry cells in the control group should remain constant over time.
I also carried out a "colour swap" version of this assay (*not illustrated above*), with the opposite combinations of fluorescent markers linked to sgRNA expression, to confirm that the fluorescent markers themselves did not affect cell fitness.

However, as previously discussed, having constitutive Cas9/sgRNA activity might select for cells that escaped or adapted to HNF1B knockout, prior to the start of each experiment. Therefore, cells expressing constitutive Cas9 were freshly transduced with sgRNAs and expanded for a minimal time period immediately prior to the start of each *in vitro* or *in vivo* experiment, to avoid selecting for cells that lack HNF1B knockout mutations or that have already adapted to HNF1B knockout before the experimental start point.

To investigate the phenotype for HNF1B knockout *in vitro*, I performed competition proliferation assays with a number of ccRCC cell lines, namely: 786-M1A, OS-LM1B, RFX631, RCC-MF, UOK101, and A498 cells (***Fig 14***). I first transduced each cell line with a constitutive Cas9 expression construct (lenti-Cas9-blast). I then lentivirally transduced each of these constitutive Cas9 cell lines with one of the following sgRNAs: the non-targeting control sgRNA #18 (NTC18, a.k.a Control), HNF1B sgRNA #2 (sg2), or HNF1B sgRNA #8 (sg8), in either an mCherry or BFP fluorescent backbone. I used the resultant cell lines (i.e. cell lines expressing Control-mCherry, Control-BFP, sg2-mCherry, sg2-BFP, sg8-mCherry, or sg8-BFP) for the subsequent proliferation assays.

The workflow for the competition assay is illustrated in ***Fig 13***. Cells expressing Control-mCherry were mixed in a 50:50 ratio with cells expressing either Control-BFP (this combination served as the control group), sg2-BFP, or sg8-BFP. To exclude the possibility that the fluorescent markers themselves could affect cell fitness, a "colour swap" version of the assay was also carried out by mixing cells expressing Control-BFP in a 50:50 ratio with cells expressing either Control-mCherry (again as the control group), sg2-mCherry, or sg8-mCherry.

These cell combinations were then cultured for several weeks, initially for 28 days (to match the time scale of the 26-day screen), and then over shorter 12 or 16 day periods as the prevailing phenotypes became apparent. The percentages of BFP and mCherry positive cells in each mixture were measured every 4 days by FACs, and used to calculate the ratio of BFP versus mCherry cells (or mCherry versus BFP cells in the case of the colour swap) over time.

If HNF1B knockout did indeed confer a selective disadvantage in ccRCC, one would expect that the sgRNA-expressing cells would either die or proliferate less than the controls, and would thus be outcompeted by the control cells in the assay mixture. Thus, over the course of the assay one would expect the percentage of sgRNA-BFP cells to decrease relative to the percentage of Control-mCherry cells (i.e. the BFP:mCherry ratio would decrease over time).

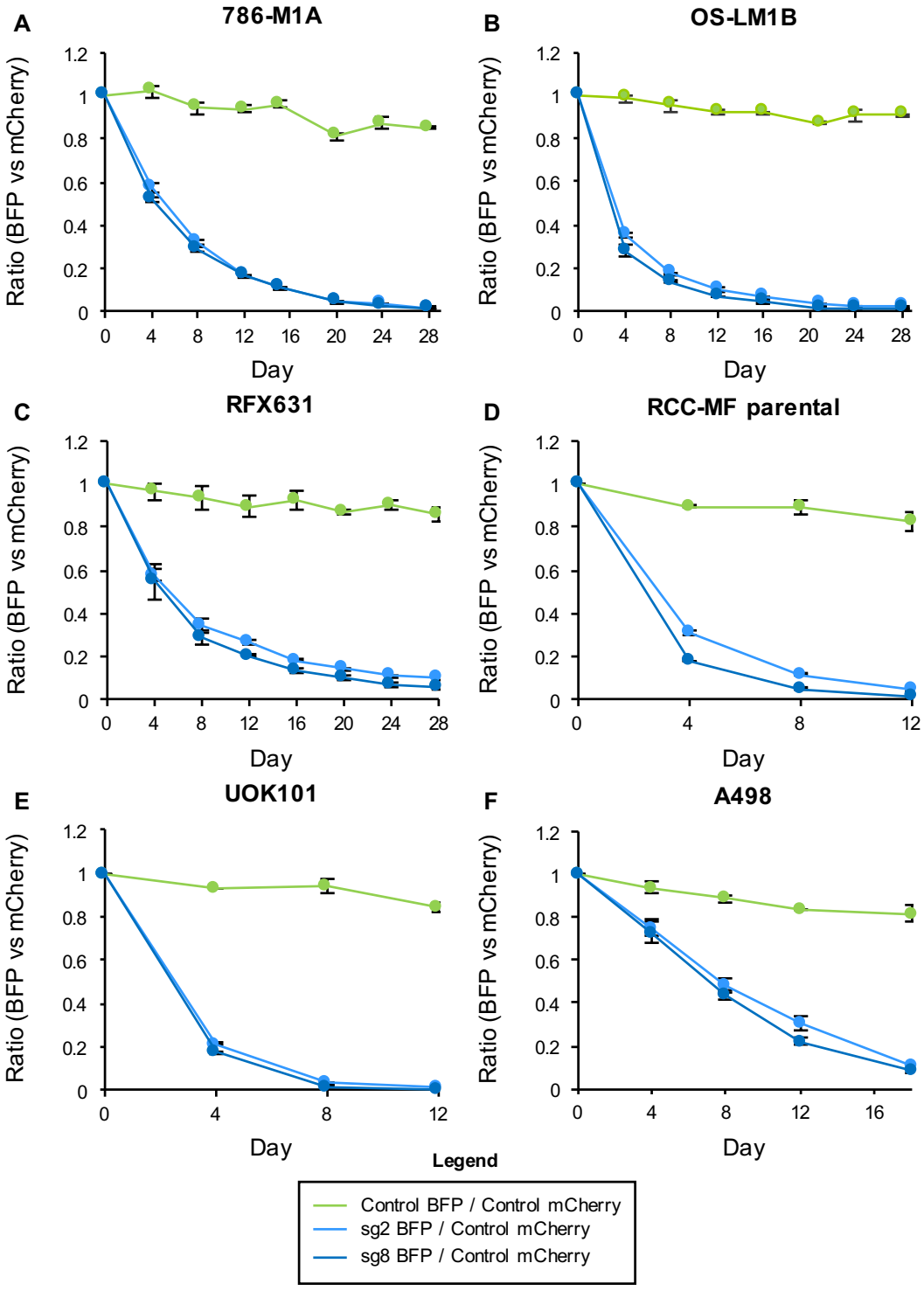


Fig 14: ***HNF1B knockout results in a striking selective disadvantage in multiple ccRCC cell lines.*** (A)-(F) Competition proliferation assay data showing the selective disadvantage conferred by HNF1B knockout in the indicated Cas9-expressing ccRCC cell lines. Control-mCherry cells were mixed in a 50:50 ratio with either Control-BFP, sg2-BFP or sg8-BFP cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the BFP:mCherry ratio over time.

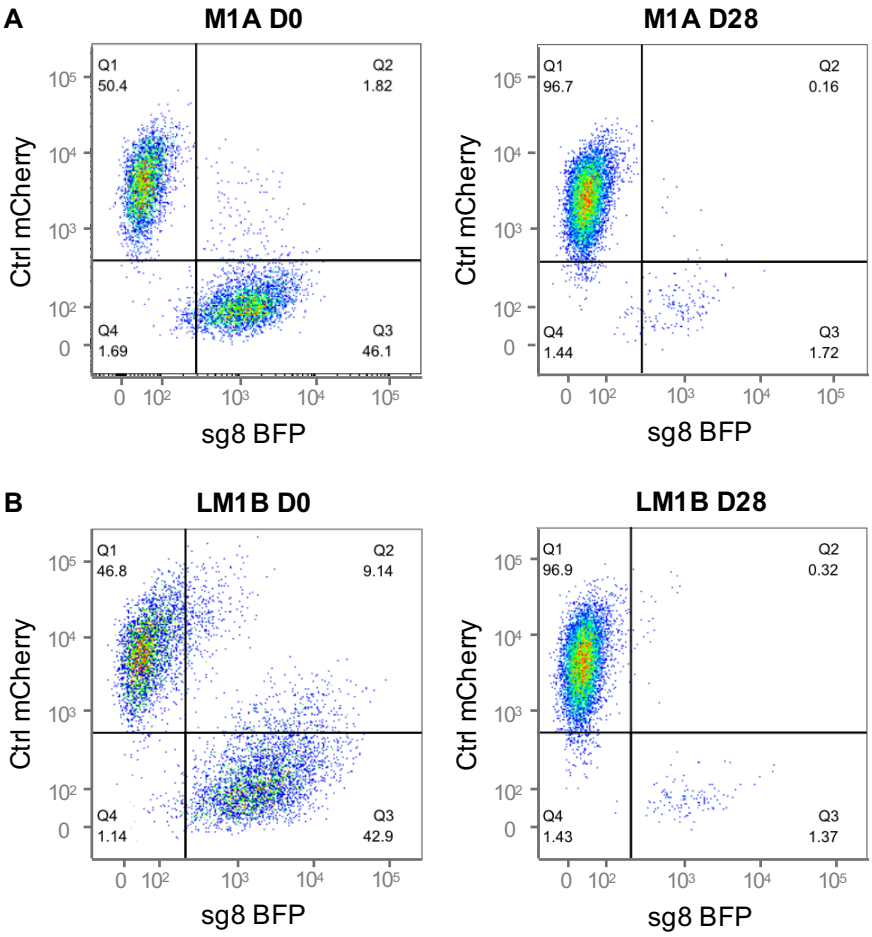


Fig 15: ***Representative FACs data from the competition assays illustrating the decrease in proportion of sgRNA-expressing cells relative to controls.*** Representative FACs data from the start (D0) and end (D28) of the competition assays performed in (A) 786-M1A-Cas9 and (B) OS-LM1B-Cas9 cells. Both sets of FACs plots show a dramatic reduction in the percentage of sgRNA-BFP cells (in this case using sg8-BFP cells as an example) relative to Control-mCherry cells in the assay mixture by the endpoint of each experiment (hence the observed decrease in the BFP:mCherry ratios over time), thus suggesting a strong selective disadvantage associated with HNF1B knockout. This data is representative of both HNF1B sgRNAs (sg2 and sg8) and all 6 cell lines used for the competition assays.

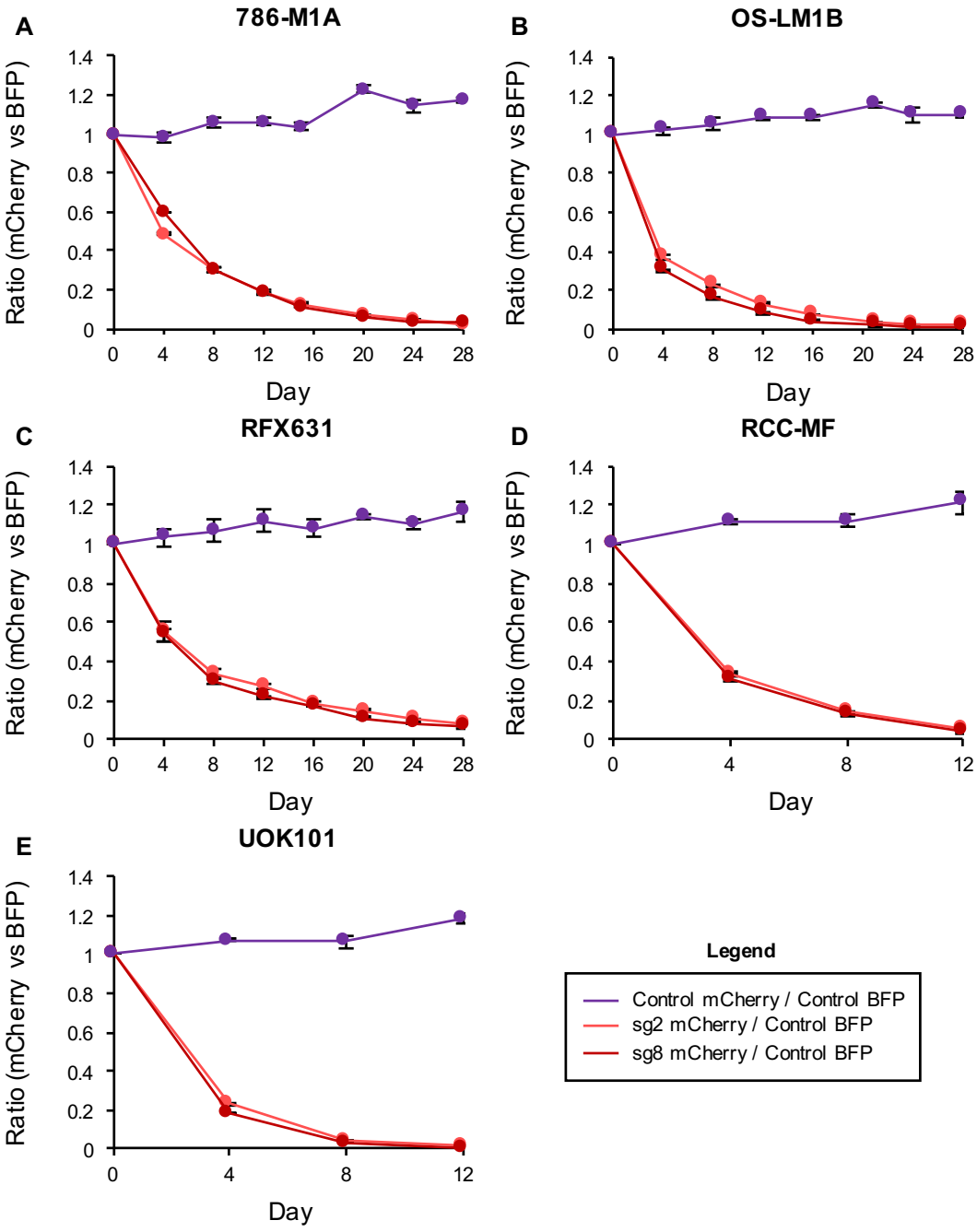


Fig 16: ***The selective disadvantage seen with HNF1B knockout is not an artefact of the combination of fluorescent markers used in the competition assay.*** (A)-(E) Proliferation data from the "colour swap" competition assays, in the indicated Cas9-expressing ccRCC cell lines. Control-BFP cells were mixed in a 50:50 ratio with either Control-mCherry, sg2-mCherry or sg8-mCherry cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the mCherry:BFP ratio over time.

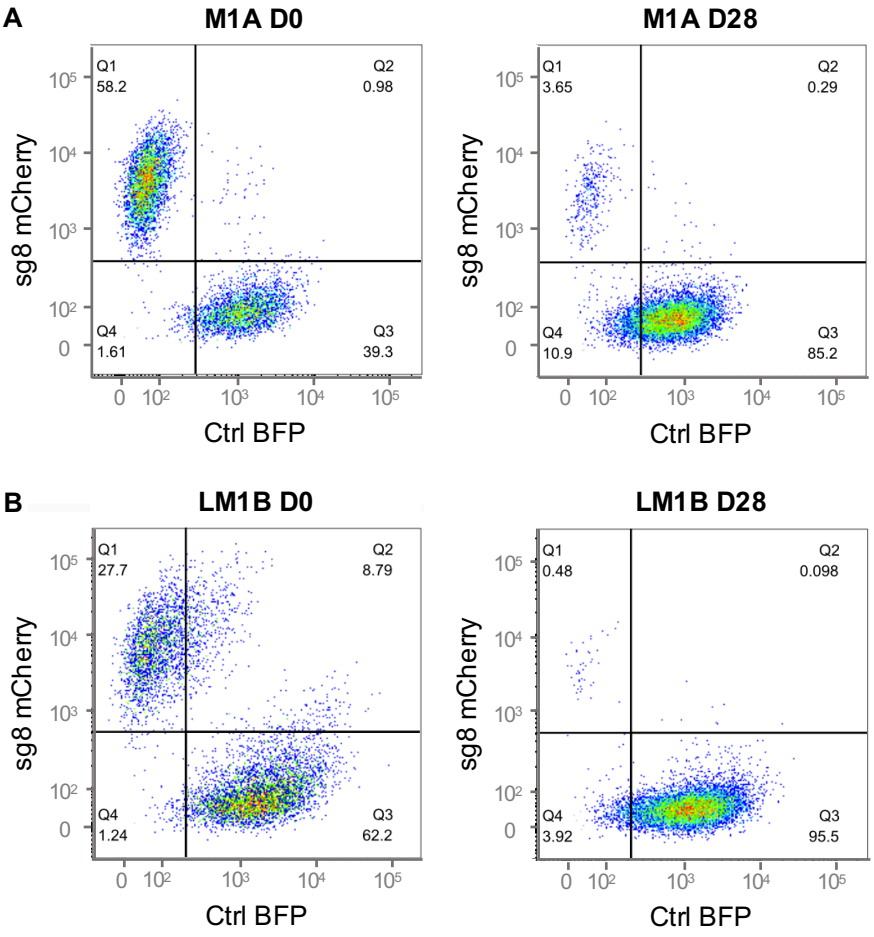


Fig 17: ***Representative FACs data from the "colour swap" competition assays illustrating the decrease in proportion of sgRNA-expressing cells relative to controls.*** Representative FACs data from the start (D0) and end (D28) of the "colour swap" competition assays performed in (A) 786-M1A-Cas9 and (B) OS-LM1B-Cas9 cells. Both sets of FACs plots show a dramatic reduction in the percentage of sgRNA-mCherry cells (in this case using sg8-mCherry cells as an example) relative to Control-BFP cells in the assay mixture by the endpoint of each experiment (hence the observed decrease in the mCherry:BFP ratios over time). Therefore, the strong selective disadvantage seen with HNF1B knockout is consistent and is not associated with a loss of cell fitness conferred by either the BFP or mCherry fluorescent markers. This data is representative of both HNF1B sgRNAs (sg2 and sg8) and all 5 cell lines used for the "colour swap" competition assays.

Similarly, in the case of the colour swap, one would expect the percentage of sgRNA-mCherry cells to decrease relative to the percentage of Control-BFP cells (i.e. the mCherry:BFP ratio would decrease over time). By contrast, the proportion of BFP versus mCherry cells in the control group (Control-mCherry cells mixed with Control-BFP cells) should remain constant over time, due to the absence of any selective pressure incurred by the non-targeting control sgRNAs. Thus, one would expect the BFP:mCherry ratio (or mCherry:BFP ratio) in the control combination to remain at approximately 1:1 over the course of the assay.

If HNF1B loss has no effect on ccRCC cells *in vitro*, then the results for the sgRNA / Control mixtures (for both the original assay and the colour swap) would be expected to resemble those of the "Control-mCherry / Control-BFP" mixture: the proportions of mCherry-positive and BFP-positive cells should remain stable over time (at approximately 1:1 ratios), as there would be no difference in selection pressure between cells expressing the non-targeting control and cells expressing HNF1B sgRNAs.

In all the cell lines tested, the BFP:mCherry ratio in the "Control-BFP / Control-mCherry" group remained constant at approximately 1:1 over time, as expected (***Fig 14A-F***). By contrast, the BFP:mCherry ratio in both the "sg2-BFP / Control-mCherry" group and the "sg8-BFP / Control-mCherry" group decreased rapidly over the course of the assay, indicating a strong selective disadvantage in the HNF1B knockout cells (as they are outcompeted by the controls). Therefore, this data demonstrates a striking phenotype for HNF1B loss *in vitro*, consistent with the strong depletion of HNF1B sgRNA abundance in the CRISPR/Cas9 screen.

This phenotype is also evident in the raw FACs data itself for all 6 ccRCC cell lines. Taking the 786-M1A-Cas9 and OS-LM1B-Cas9 Control and sg8 cell lines as representative examples (***Fig 15A-B***), if we compare the FACs plots of BFP against mCherry fluorescence on Day 0 with those on Day 28 for each competition assay we can see a dramatic reduction in the percentage of sgRNA-BFP cells relative to Control-mCherry cells in the assay mixture by the endpoint of each experiment (hence the observed decrease in the BFP:mCherry ratio over time), thus suggesting a strong selective disadvantage associated with HNF1B knockout. The data shown in ***Fig 15A-B*** is representative of both sg2 and sg8, across all 6 cell lines.

Similarly, in the "colour swap" competition assays, the mCherry:BFP ratio in the "Control-mCherry / Control-BFP" group remained at approximately 1:1, while the mCherry:BFP ratio in both "sgRNA-mCherry / Control-BFP" groups decreased rapidly over time, in all the cell

lines tested (***Fig 16A-E***). Thus, the same strong selective disadvantage is observed with HNF1B knockout, irrespective of the different fluorescent marker linked to HNF1B sgRNA expression. This verifies that the phenotype seen with HNF1B depletion is unlikely to be the result of a loss of cell fitness associated with either the BFP or mCherry fluorescent markers.

Using the 786-M1A-Cas9 and OS-LM1B-Cas9 Control and sg8 cells as representative examples once again, FACs plots from the start (Day 0) and end (Day 28) of the "colour swap" competition assays also show a dramatic reduction in the percentage of sgRNA-mCherry cells relative to Control-BFP cells in the assay mixture by the endpoint of each experiment (hence the observed decrease in the mCherry:BFP ratio over time) (***Fig 17A-B***). This further emphasises that the strong selective disadvantage seen with HNF1B knockout is reproducible, and is not associated with a loss of cell fitness conferred by either the BFP or mCherry fluorescent markers. The data shown in ***Fig 17A-B*** is representative of both sg2 and sg8, across all 5 cell lines.

To ensure that the selective disadvantage seen with HNF1B knockout in the competition assays was not an artefact resulting from mixing HNF1B wild-type and deficient cells, I also investigated the phenotype using more conventional in vitro proliferation assays. Specifically, I performed trypan blue cell counting and CellTitre Glo proliferation assays with 786-M1A-Cas9 and OS-LM1B-Cas9 cells expressing either Control-BFP, sg2-BFP, or sg8-BFP. I used the same batch of transduced cell lines as those used for the competition assays (***Figs 14A-B, 15A-B***), and seeded the cell counting and CellTitre Glo experiments at roughly the same time as the competition assays so that the cells in all three experiments would have experienced HNF1B depletion for the same number of days. Otherwise, some experiments may have contained knockout cell lines that had progressed further towards selecting for subclones that had escaped or become resistant to HNF1B depletion.

In both ccRCC cell line systems, and for both the cell counting (***Fig 18A-B***) and CellTitre Glo (***Fig 18C-D***) assays, there is a clear selective disadvantage seen in the cells expressing HNF1B sg2 and sg8, relative to the control cells. This is consistent with, and therefore validates, the phenotype seen in the competition assays (where HNF1B knockout cells were outcompeted by the controls).

I also carried out cell counting and CellTitre Glo proliferation assays with 786-M1A-Cas9 and OS-LM1B-Cas9 cells expressing either Control-mCherry, sg2-mCherry, or sg8-mCherry. Once again, I used the same batch of transduced cell lines as those used for the

competition assays (***Figs 16A-B, 17A-B***), and seeded all three proliferation experiments (the cell counting, CellTitre Glo and competition assays) at the same time to avoid using cells at different stages of knockout selection pressure.

As seen previously, cells expressing HNF1B sg2 and sg8 proliferate at a markedly reduced rate compared to the control cells (***Fig 19A-D***). This phenotype was seen in both ccRCC cell line systems and for both the cell counting (***Fig 19A-B***) and CellTitre Glo (***Fig 19C-D***) assays. Moreover, this data is consistent with the data obtained from the BFP cell lines, and from the competition assays. Thus, this further establishes that the selective disadvantage observed following HNF1B loss is reproducible across multiple assay types, and is independent of the fluorescent marker linked to sgRNA expression.

Futhermore, both the cell counting and CellTitre Glo assays do not involve mixing control and knockout cells, thus avoiding the potential effects of cell signalling between cell lines that could influence cell proliferation. Therefore, the phenotype seen with HNF1B knockout is not a function of mixing cell lines, as the same phenotype is observed in more conventional proliferation assays.

To confirm that HNF1B was successfully depleted in the competition, cell counting, and CellTitre Glo proliferation assays, I cultured the individual BFP and mCherry cell lines independently (in parallel with the cell line combinations in each competition assay), and used these "solo" cell line cultures to obtain cell pellets for protein and/or RNA samples on Day 0 and at each subsequent 4-day assay time point. I then extracted whole cell protein lysates from the Day 0 pellets and blotted for HNF1B in the BFP versions (***Fig 20A***) and mCherry versions (***Fig 20B***) of each cell line, to confirm that cells expressing sg2 or sg8 were HNF1B-deficient, and cells expressing the non-targeting control sgRNA were still HNF1B wild-type, at the start of each assay.

Indeed, in all the ccRCC cell line systems tested, HNF1B was strongly expressed in cells transduced with the non-targeting control sgRNA. By contrast, cells transduced with HNF1B sg2 or sg8 showed a highly efficient knockout with an almost complete depletion of HNF1B protein (***Fig 20A-B***).

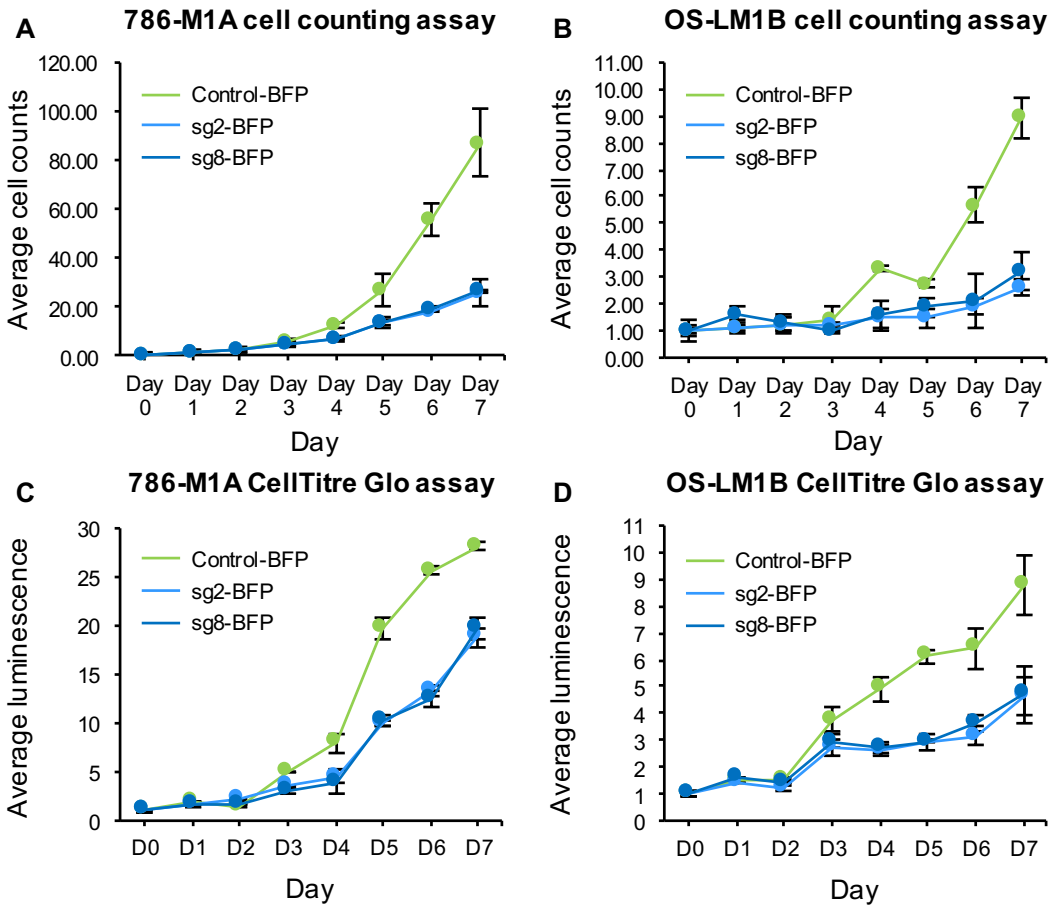


Fig 18: ***The selective disadvantage associated with HNF1B knockout is also observed in more conventional in vitro proliferation assays.*** (A)-(B) Trypan blue cell counting proliferation assays and (C)-(D) CellTitre Glo proliferation assays, performed with 786-M1A-Cas9 and OS-LM1B-Cas9 cells transduced with Control-BFP, sg2-BFP, or sg8-BFP. Consistent with the data from the competition assays, the HNF1B knockout (sg2 and sg8) cells exhibited a proliferative disadvantage relative to the Control cells. This reduced proliferation was observed in both ccRCC cell line systems, and in both types of *in vitro* proliferation assay (neither of which involve mixing control and knockout cells), thus validating the data obtained from the competition assays.

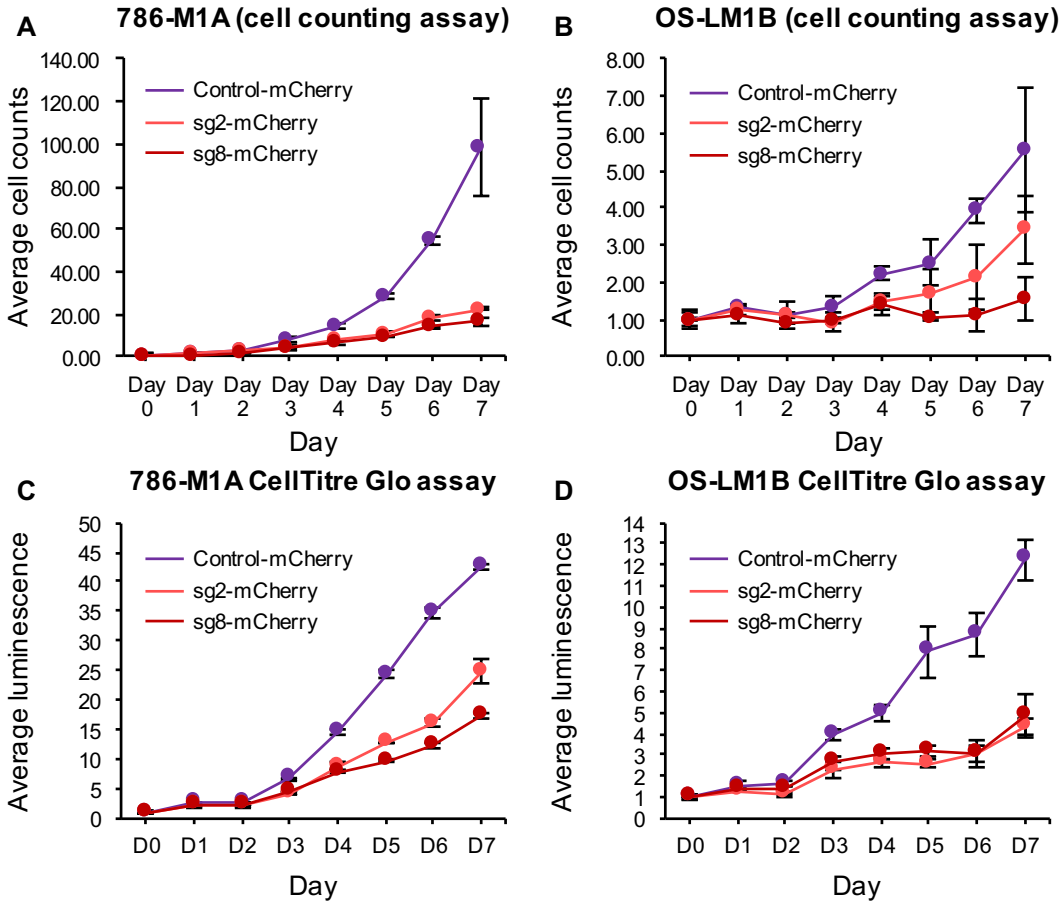


Fig 19: ***The selective disadvantage seen with HNF1B knockout in conventional in vitro proliferation assays is consistent and reproducible between BFP and mCherry cell lines.*** (A)-(B) Trypan blue cell counting proliferation assays and (C)-(D) CellTitre Glo proliferation assays performed with 786-M1A-Cas9 and OS-LM1B-Cas9 cells, this time transduced with Control-mCherry, sg2-mCherry, or sg8-mCherry. Once again, consistent with the data from the competition assays, the HNF1B knockout (sg2 and sg8) cells exhibited a proliferative disadvantage relative to the Control cells. This was observed in both ccRCC cell line systems and in both types of *in vitro* proliferation assay (neither of which involve mixing control and knockout cells). This validates the data from the competition assays and the data from the Trypan blue cell counting CellTitre Glo proliferation assays performed with the Control-, sg2-, and sg8-BFP cell lines.

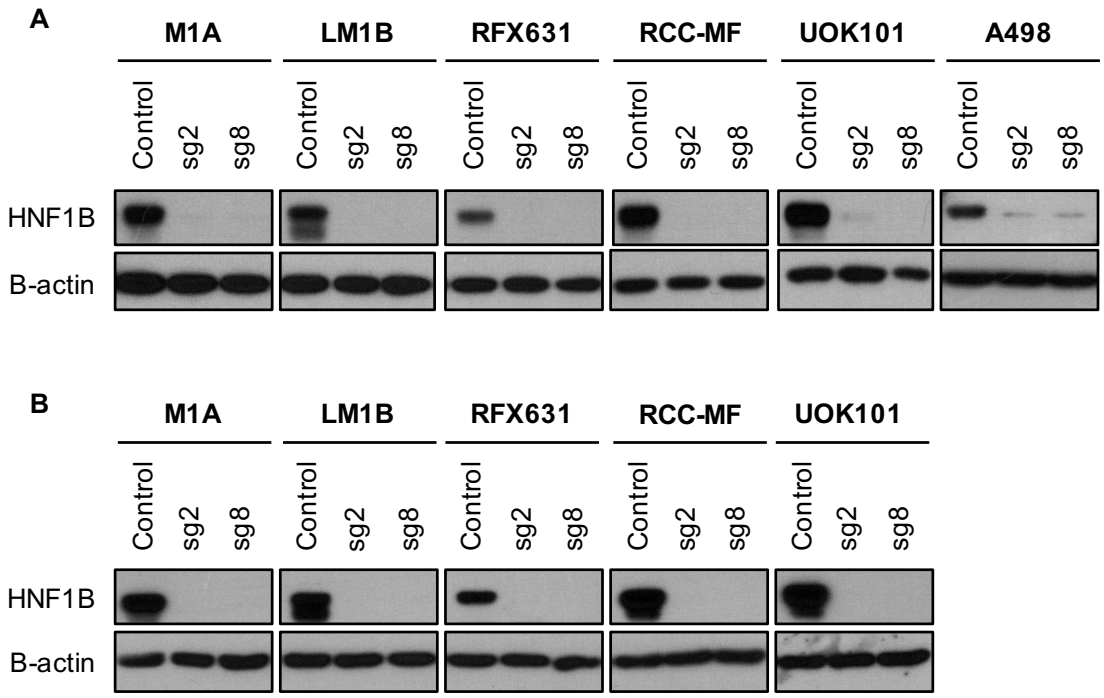


Fig 20: ***Western blotting for HNF1B in the cell lines used for the proliferation and competition assays.*** Western blots for HNF1B in whole cell protein lysates extracted on Day 0 from (A) the BFP versions and (B) the mCherry versions of the indicated ccRCC cell lines, which were used for the competition, cell counting and CellTitre Glo proliferation assays.

2.3) The selective disadvantage with HNF1B knockout does not seem to be an off-target effect

To confirm that the selective disadvantage seen with HNF1B loss was not the result of an off-target effect of the HNF1B sgRNAs, I also performed the same competition proliferation assays (including the "colour swap") in HeLa cells, which do not express HNF1B. One would expect that HNF1B knockout would have no effect in a cell line that does not express HNF1B, unless the HNF1B sgRNAs induced an off-target effect that adversely affected cell survival or proliferation.

I first transduced parental, unmodified HeLa cells with the same constitutive Cas9 expression construct (lenti-Cas9-blast) used in the ccRCC cells. I then lentivirally transduced the HeLa-Cas9 cells with the same sgRNA constructs used for the ccRCC competition assays (i.e. Control, HNF1B sg2, or HNF1B sg8, each in either an mCherry or BFP backbone).

As before, I mixed Control-mCherry cells in a 50:50 ratio with cells containing either Control-BFP (the control group), sg2-BFP, or sg8-BFP. Meanwhile, for the "colour swap", I mixed Control-BFP cells in a 50:50 ratio with cells containing either Control-mCherry (again as the control group), sg2-mCherry, or sg8-mCherry. As with the ccRCC Cas9 competition assays, I cultured these cell combinations for 28 days while measuring the percentages of BFP and mCherry positive cells in each mixture every 4 days by FACs, and calculated the ratio of BFP versus mCherry cells (or mCherry versus BFP cells for the "colour swap") over time.

The BFP:mCherry ratio in the "Control-BFP / Control-mCherry" group (***Fig 21A***), and the mCherry:BFP ratio in the "Control-mCherry / Control-BFP" group for the "colour swap" (***Fig 21B***), both remained at approximately 1:1 as expected. However, in contrast to the ccRCC competition assays, the BFP:mCherry ratio in the "sg2-BFP / Control-mCherry" and "sg8-BFP / Control-mCherry" groups stayed relatively constant, at a roughly 1:1 ratio over time (***Fig 21A***). Similarly, in the "colour swap" assay, the mCherry:BFP ratio in both "sgRNA-mCherry / Control-BFP" groups also remained at approximately 1:1, again phenocopying the control group (and again indicating that there is no loss of cell fitness associated with either the BFP or mCherry fluorescent markers) (***Fig 21B***). In other words, HeLa-Cas9 cells expressing sg2 or sg8 seemed to proliferate at the same rate as cells expressing the non-targeting control, and did not exhibit the selective disadvantage seen in ccRCC cells expressing HNF1B sgRNAs.

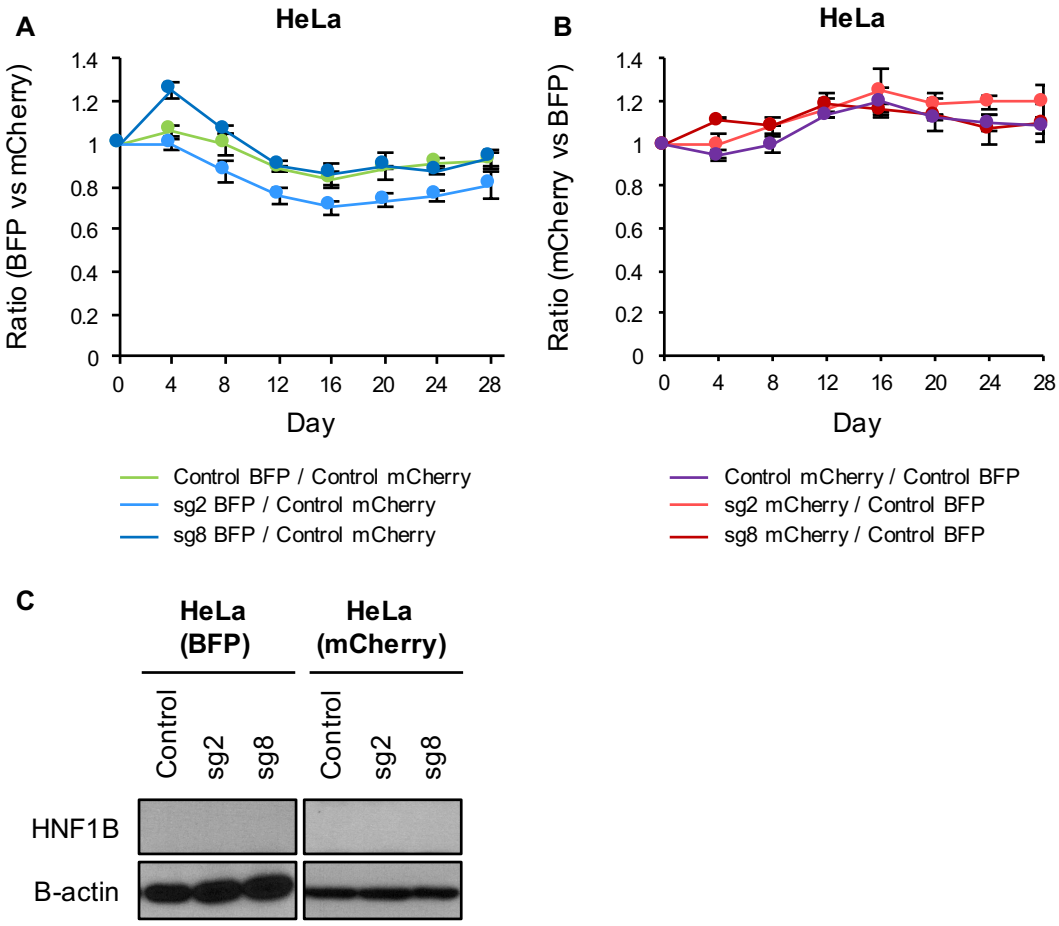


Fig 21: ***HNF1B knockout has no effect in non-ccRCC cell lines that do not express HNF1B.*** Proliferation data from the competition assays performed in HeLa-Cas9 cells, for both (A) the original fluorescent marker combination and (B) the "colour swap" combination. (C) Western blots for HNF1B in whole cell protein lysates extracted on Day 0 from the BFP versions (*left*) and the mCherry versions (*right*) of the HeLa-Cas9 cell lines used for the competition proliferation assays.

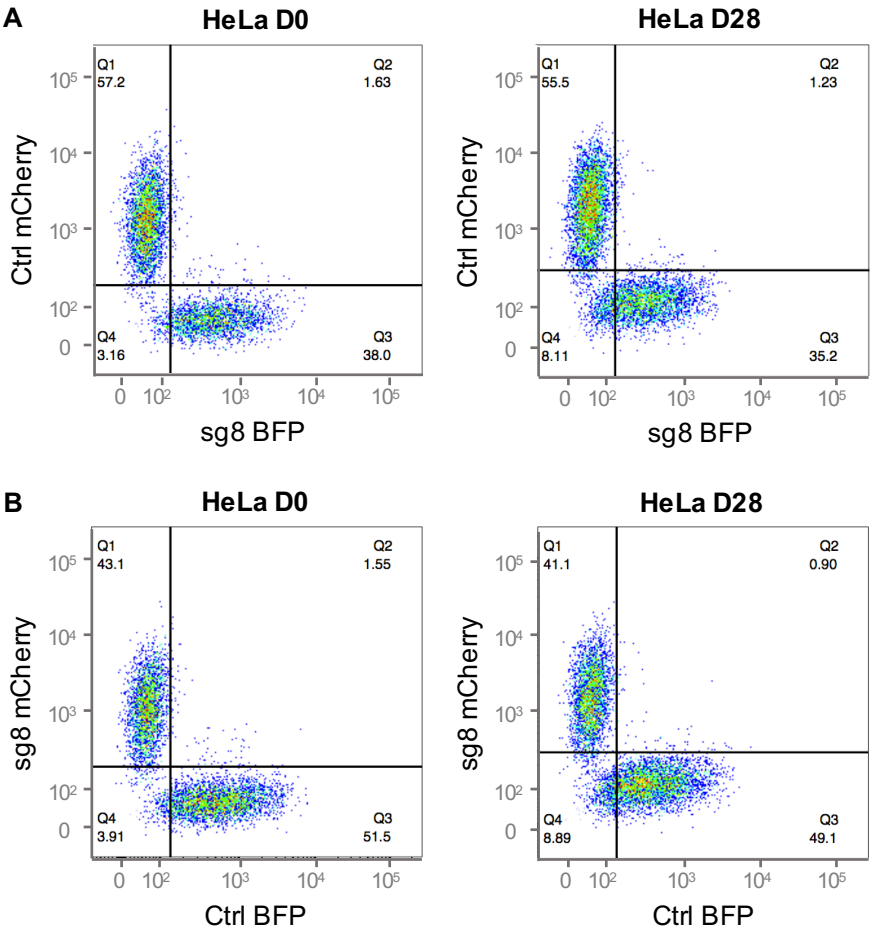


Fig 22: ***Representative FACs data from the competition assays in HeLa-Cas9 cells, illustrating the lack of a phenotype for HNF1B knockout in these cells.***
Representative FACs data from the start (D0) and end (D28) of the competition assays performed in HeLa-Cas9 cells, for both (A) the original fluorescent marker combination and (B) the "colour swap" combination. Both sets of FACs plots show no observable change in the percentage of sgRNA-expressing cells (in this case using sg8-BFP/mCherry cells as an example) relative to Control cells in the assay mixture by the endpoint of each experiment, regardless of the fluorescent marker associated (hence the stable 1:1 BFP:mCherry and mCherry:BFP ratios over time). This data is representative of both HNF1B sgRNAs (sg2 and sg8).

This lack of a phenotype in a cell line that does not express HNF1B suggests that the selective disadvantage observed with HNF1B knockout is not the result of an off-target effect of the HNF1B sgRNAs, but is instead associated with the loss of HNF1B protein.

To verify that the HeLa-Cas9 Control- and sgRNA-expressing cell lines were indeed HNF1B-deficient, I performed Western blots for HNF1B in whole cell protein lysates from the cells used in the competition assay. As with the ccRCC cell lines, I cultured the individual HeLa-Cas9 Control, sg2, and sg8 BFP and mCherry cell lines in parallel with the competition assay cell mixtures, and took cell pellets for protein and/or RNA on Day 0 and at each time point. I then extracted whole cell protein lysates from the Day 0 pellets and blotted for HNF1B in all 6 cell lines. The subsequent Western blot data confirmed that the HeLa cell lines used in the competition assays were all HNF1B-deficient (***Fig 21C***).

As before, the lack of a phenotype in HeLa-Cas9 cells is also evident in the raw FACs data from the start (D0) and end (D28) of the competition assays performed, for both the original fluorescent marker combination (***Fig 22A***) and the "colour swap" combination (***Fig 22B***). Taking the HeLa-Cas9 Control and sg8 cells as examples, both sets of FACs plots show no observable change in the percentage of sgRNA-expressing cells relative to Control cells in the assay mixture by the endpoint of each experiment, regardless of the fluorescent marker associated (hence the stable 1:1 BFP:mCherry and mCherry:BFP ratios over time). Note that the data shown in Fig ***22A-B*** is representative of both sg2 and sg8.

This further demonstrates that there is no phenotype associated with HNF1B sgRNA expression in HeLa cells (which do not express HNF1B), and thus supports the conclusion that the selective disadvantage observed with HNF1B knockout in ccRCC cell lines is most likely the result of HNF1B depletion rather than an off-target effect of the sgRNAs used.

2.4) The selective disadvantage with HNF1B knockout seems to be specific to HNF1B loss

To confirm that the phenotype seen with HNF1B knockout was indeed specific to the loss of HNF1B protein, I performed competition proliferation assays comparing control cells (transduced with the Control sgRNA NTC18, and empty vector), HNF1B knockout cells without rescue (transduced with HNF1B sg8, and empty vector), and HNF1B knockout cells *with* rescue (transduced with HNF1B sg8, and an exogenous HNF1B rescue construct comprising an exogenous HNF1B transgene under the control of a constitutive promoter, with the HNF1B

cDNA sequence modified by site-directed mutagenesis to mutate the sg8 target site and make it untargetable, thus preventing Cas9 cleavage of the transgene cDNA).

786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cells were first transduced with either empty vector or the exogenous HNF1B rescue construct (constitutively expressed), and then transduced with either Control-BFP, Control-mCherry, sg8-BFP or sg8-mCherry. The resultant cell lines were subsequently used for the competition assay mixtures. Note that cell lines containing the rescue construct are denoted "(+ rescue)" in ***Figure 23***; all other cell lines contain only the empty vector.

For the first three experimental groups, Control-mCherry (+ empty vector) cells were mixed in a 50:50 ratio with either Control-BFP (+ empty vector), sg8-BFP (+ empty vector) or sg8-BFP (+ rescue) cells. In a fourth experimental group, sg8-mCherry (+ rescue) cells were mixed in a 50:50 ratio with sg8-BFP (+ empty vector) cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the BFP:mCherry ratio over time.

Consistent with previous results, the BFP:mCherry ratio remained at approximately 1:1 in the "Control-BFP / Control-mCherry" group, but decreased rapidly in the "sg8 BFP / Control mCherry" group, in both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell lines (***Fig 23A-B***).

The "sg8 (+ rescue) BFP / Control mCherry" group, however, exhibited a phenotype more similar to the Control group, with the BFP:mCherry ratio staying relatively high and closer to 1:1 over time (***Fig 23A-B***). This suggests that knockout cells expressing the exogenous HNF1B rescue construct experience a recovery of cell fitness, enabling them to survive and proliferate just as well as HNF1B wild-type control cells. Thus, reintroducing HNF1B expression in cells that have lost the endogenous protein seems to rescue the selective disadvantage seen with HNF1B knockout.

Consistent with this, the "sg8 BFP / sg8 (+ rescue) mCherry" group phenocopied the "sg8 BFP / Control mCherry" group, with the BFP:mCherry ratio also falling dramatically over time (***Fig 23A-B***). In other words, knockout cells expressing the exogenous HNF1B transgene are able to outcompete knockout cells without rescue, with a similar efficiency to control cells. This further suggests that HNF1B restoration serves to rescue the loss of cell fitness in knockout cells, allowing them to survive and proliferate just as well as cell lines expressing a non-targeting sgRNA.

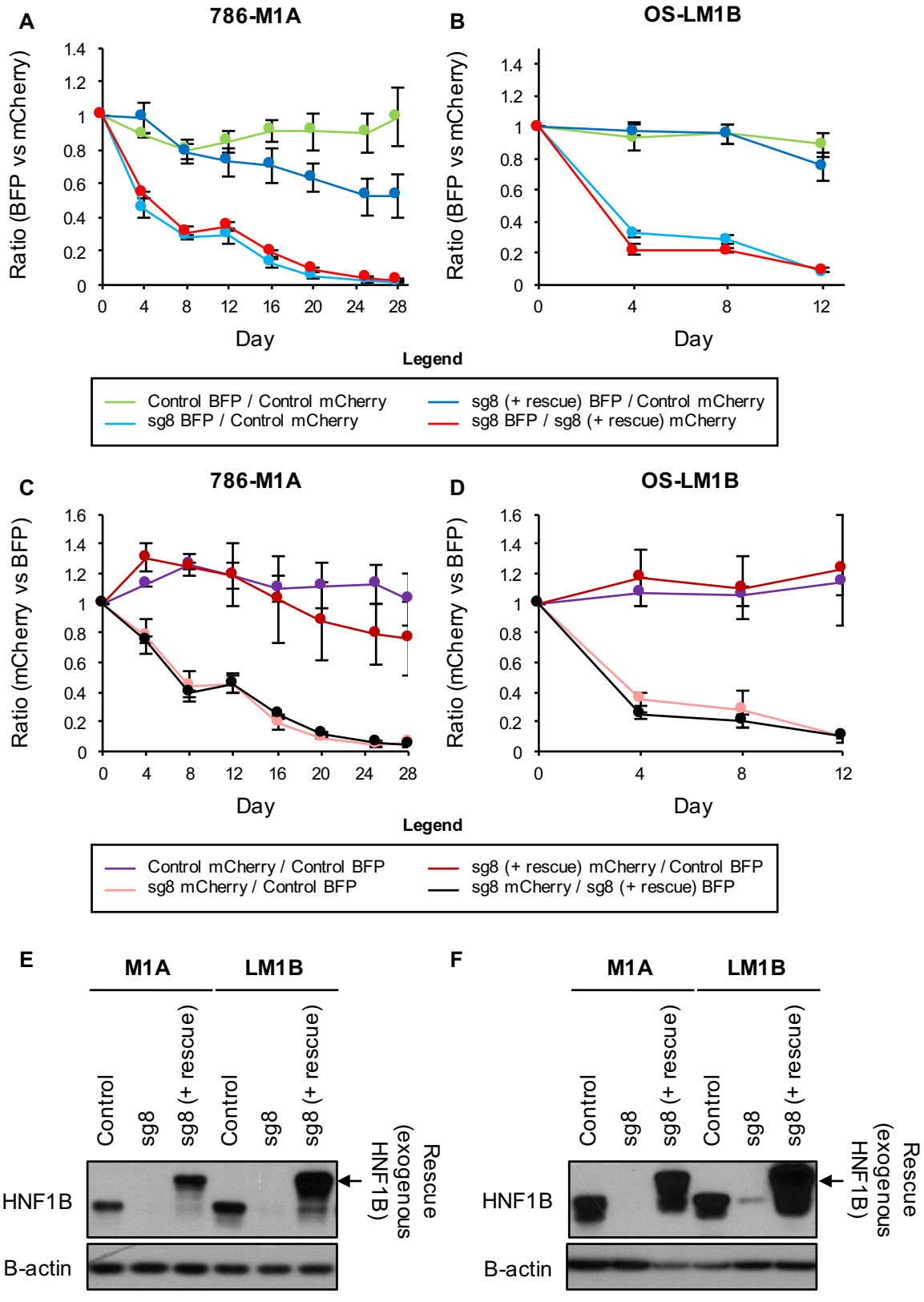


Fig 23: ***The selective disadvantage associated with HNF1B knockout is rescued by exogenous HNF1B restoration.*** P.T.O for figure legend

Fig 23: ***The selective disadvantage associated with HNF1B knockout is rescued by exogenous HNF1B restoration.***
(A)-(B) Proliferation data from the rescue competition assays performed with 786-M1A-Cas9 (*A*) and OS-LM1B-Cas9 (*B*) cells. In the first three experimental groups, Control-mCherry (+ empty vector) cells were mixed in a 50:50 ratio with either Control-BFP (+ empty vector), sg8-BFP (+ empty vector) or sg8-BFP (+ rescue) cells. In a fourth experimental group, sg8-mCherry (+ rescue) cells were mixed in a 50:50 ratio with sg8-BFP (+ empty vector) cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the BFP:mCherry ratio over time.

(C)-(D) Proliferation data from the "colour swap" rescue competition assays performed with 786-M1A-Cas9 (*C*) and OS-LM1B-Cas9 (*D*) cells. In the first three experimental groups, Control-BFP (+ empty vector) cells were mixed in a 50:50 ratio with either Control-mCherry (+ empty vector), sg8-mCherry (+ empty vector) or sg8-mCherry (+ rescue) cells. In the fourth experimental group, sg8-BFP (+ rescue) cells were mixed in a 50:50 ratio with sg8-mCherry (+ empty vector) cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the mCherry:BFP ratio over time.

(E)-(F) Western blots for HNF1B in whole cell protein lysates extracted on Day 0 from the (*E*) BFP versions and the (*F*) mCherry versions of the indicated 786-M1A-Cas9 and OS-LM1B-Cas9 cell lines, which were used for the rescue competition proliferation assays. Endogenous HNF1B appears at approximately 61kDa, while exogenous HNF1B appears at 72kDa (owing to the presence of a 6xMyc tag at the N-terminus of the protein).

Taken together, this data suggests that the phenotype observed with HNF1B knockout is most likely specifically associated with HNF1B depletion, rather than being the result of an off-target effect.

The "colour swap" versions of the rescue competition assays showed similar results. In the first three experimental groups, Control-BFP (+ empty vector) cells were mixed in a 50:50 ratio with either Control-mCherry (+ empty vector), sg8-mCherry (+ empty vector) or sg8-mCherry (+ rescue) cells. In the fourth experimental group, sg8-BFP (+ rescue) cells were mixed in a 50:50 ratio with sg8-mCherry (+ empty vector) cells. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and used to calculate the mCherry:BFP ratio over time.

As seen in previous experiments, the mCherry:BFP ratio remained at approximately 1:1 in the "Control mCherry / Control BFP" group, but decreased rapidly in the "sg8 mCherry / Control BFP" group, in both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell lines (***Fig 23C-D***).

The "sg8 (+ rescue) mCherry / Control BFP" group again exhibited a phenotype more similar to the Control group, with the mCherry:BFP ratio staying relatively high and closer to 1:1 over time (***Fig 23C-D***). Meanwhile, as seen before, the "sg8 mCherry / sg8 (+ rescue) BFP" group phenocopied the "sg8 mCherry / Control BFP" group, with the mCherry:BFP ratio decreasing similarly over time (***Fig 23C-D***).

Therefore, this data supports the idea that HNF1B rescue serves to restore cell fitness in knockout cells, enabling them to survive and proliferate to a similar extent as HNF1B wild-type controls. Additionally, as with the other "colour swap" competition assays, this set of experiments also establishes that the selective disadvantage seen with HNF1B knockout, and the phenotypic rescue seen with HNF1B restoration, are observable regardless of the fluorescent marker being expressed in the cells, and thus that these phenomena are not artefacts of the assay itself.

During each rescue competition assay, I also cultured the individual BFP and mCherry cell lines independently, in parallel with the cell line combinations in each experiment. As with the previous competition assays, I used these "solo" cell line cultures to obtain cell pellets for protein and/or RNA samples on Day 0, and at each subsequent 4-day assay time point. I then extracted whole cell protein lysates from the Day 0 pellets for each cell line (both BFP and mCherry versions), and blotted for HNF1B.

The Western blot data confirmed that, at the start of each assay, the Control (+ empty vector) cells still expressed endogenous wild-type HNF1B, while the HNF1B knockout cells without rescue (transduced with sg8 + empty vector) were HNF1B-deficient. Meanwhile, the HNF1B knockout cells *with* rescue (transduced with sg8 + the rescue construct) expressed considerably less endogenous HNF1B, while simultaneously expressing high levels of exogenous HNF1B protein (***Fig 23E-F***). Note that endogenous HNF1B appears at approximately 61kDa, while exogenous HNF1B appears at 72kDa owing to the presence of a 6xMyc tag at the N-terminus of the protein.

Therefore, this data verifies that in both rescue competition assays, the non-targeting control sgRNA had no effect on HNF1B expression, sg8 had successfully mediated efficient constitutive knockout of only endogenous HNF1B protein, and the exogenous HNF1B transgene was successfully expressed at levels that are similar to, or higher than, the levels of endogenous HNF1B protein.

Even if the expression levels of exogenous HNF1B exceeded those of endogenous HNF1B, the extent of rescue was similar at the Day 12 time point between both sets of 786-M1A-Cas9 and OS-LM1B-Cas9 rescue competition assays (***Fig 23A-D***), thus suggesting that the exogenous HNF1B transgene was expressed without reaching levels of overexpression that would introduce artifacts in the data.

2.5) Investigating the phenotype for HNF1B CRISPRi knockdown in vitro

To confirm that the selective disadvantage seen with HNF1B CRISPR/Cas9 knockout was not an effect of double-stranded break formation by Cas9, I also performed competition proliferation assays involving HNF1B knock*down*, using the CRISPRi system.

The CRISPRi system employs the use of a catalytically dead Cas9 protein (dCas9) fused to a KRAB domain. The dCas9 protein is targeted to specific genomic DNA sites by sgRNAs in the same manner as wild-type Cas9. However, instead of creating a double-stranded break (DSB), the dCas9 protein binds to the DNA and the attached KRAB domain mediates transcriptional repression at that site[125,126].

I used 786-M1A and OS-LM1B ccRCC cell lines in the lab that stably expressed a dCas9-KRAB-mCherry transgene. I designed a number of tandem HNF1B sgRNA constructs to mediate protein knockdown with the CRISPRi system. Each tandem construct comprised two

HNF1B sgRNAs (specifically designed for CRISPRi), each flanked by its own U6 promoter and sgRNA scaffold. In other words, the tandem construct structure (devised by a postdoc in the lab), was as follows: U6 promoter—HNF1B CRISPRi sgRNA—sgRNA scaffold—U6 promoter—HNF1B CRISPRi sgRNA—sgRNA scaffold. The rationale behind using two sgRNAs instead of one was that this approach reportedly promotes better targeting efficiency and specificity[127] for the dCas9 enzyme. The CRISPRi sgRNA sequences were obtained from either the genome-scale CRISPRi sgRNA library designed by Gilbert *et al*.[126], or from the Broad Institute online sgRNA design tool (https://portals.broadinstitute.org/gpp/public/).

I selected 3 tandem HNF1B sgRNA constructs for use: two constructs that were designed to target HNF1B but ultimately did not alter HNF1B levels ("HNF1B CRISPRi #1" and "HNF1B CRISPRi #3", both of which acted as controls), and a third tandem construct that successfully mediated HNF1B knockdown ("HNF1B CRISPRi #2"). As a control, I used a tandem construct comprised of two non-targeting "scramble" CRISPRi sgRNAs (the "Scramble Control") already available in the lab.

All four tandem constructs were cloned into a BFP fluorescent backbone. I transduced them into 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells, and sorted for the brightest mCherry+/BFP+ population to isolate successfully transduced cells expressing high levels of the dCas9-KRAB and tandem sgRNA constructs. However, with mCherry already expressed in the cell lines (originally designed as a transduction selection marker and to provide a readout of dCas9-KRAB expression), I had to modify the competition assay set-up.

Thus, instead of comparing the proportions of BFP-expressing vs mCherry-expressing cells in my assay mixtures, I mixed untransduced parental BFP– cells (not transduced with sgRNAs) in a 50:50 ratio with BFP+ cells transduced with either the Scramble Control, HNF1B CRISPRi #1, HNF1B CRISPRi #2 or HNF1B CRISPRi #3 constructs. I then measured the percentages of BFP positive versus BFP negative cells every 4 days by FACs, and calculated the BFP+/BFP– ratio over time.

The BFP+/BFP– ratio in the "HNF1B CRISPRi #1 vs untransduced" and "HNF1B CRISPRi #3 vs untransduced" experimental groups remained relatively high (closer to a 1:1 ratio) and stable over the course of the assay in both cell lines (***Fig 24A-B***). This is consistent with the fact that neither of these constructs successfully facilitate HNF1B knockdown (***Fig 24C-***

D). By contrast, the BFP+/BFP– ratio in the "HNF1B CRISPRi #2 vs untransduced" groups decreased rapidly over time in both cell lines (***Fig 24A-B***), similar to the striking phenotype seen with HNF1B knockout versus control cells in the competition assays performed with the CRISPR/Cas9 system (***Fig 14A-B***, ***Fig 16A-B***). Indeed, the HNF1B CRISPRi #2 tandem sgRNA construct is the only one that successfully knocks down HNF1B (***Fig 24C-D***). This indicates that cells experiencing HNF1B knockdown are rapidly outcompeted by untransduced HNF1B wild-type cells in the mixture, and are thus under a strong selective disadvantage.

Therefore, this data reaffirms the striking phenotype for HNF1B loss *in vitro*. Furthermore, the consistent phenotypes observed with HNF1B CRISPRi knockdown (which is based on transcriptional repression and does not involve the formation of DSBs) and HNF1B CRISPR/Cas9 knockout (which does involve the formation of DSBs) suggest that the selective disadvantage associated with HNF1B depletion is indeed related to the loss of HNF1B and its functions, rather than a loss of cell fitness resulting from double-stranded break formation by Cas9.

To determine whether each HNF1B tandem sgRNA construct was capable of knocking down HNF1B, I extracted whole cell protein lysates and total RNA from the cell lines used in each CRISPRi competition assay, and performed qPCRs (***Fig 24C***) and Western blots (***Fig 24D***) for HNF1B.

For the qPCR data in the 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry sets of cell lines, I normalised the HNF1B mRNA levels in each HNF1B tandem sgRNA sample to the HNF1B mRNA levels in the corresponding Scramble Control sample for each set of cell lines, to check for any relative decreases in HNF1B expression. The HNF1B CRISPRi #2 construct mediated an 88.9% decrease (a fold change of 0.111) in HNF1B mRNA levels in 786-M1A-dCas9-KRAB-mCherry cells, and a 52% decrease (a fold change of 0.480) in OS-LM1B-dCas9-KRAB-mCherry cells (***Fig 24C***). By contrast, the HNF1B CRISPRi #1 and HNF1B CRISPRi #3 constructs failed to knock down HNF1B in both cell line systems (***Fig 24C***).

This is also reflected in the Western blot data, where HNF1B protein levels are reduced in cells expressing HNF1B CRISPRi #2 compared to cells expressing the Scramble Control (***Fig 24D***). However, cells expressing HNF1B CRISPRi #1 and HNF1B CRISPRi #3 show HNF1B

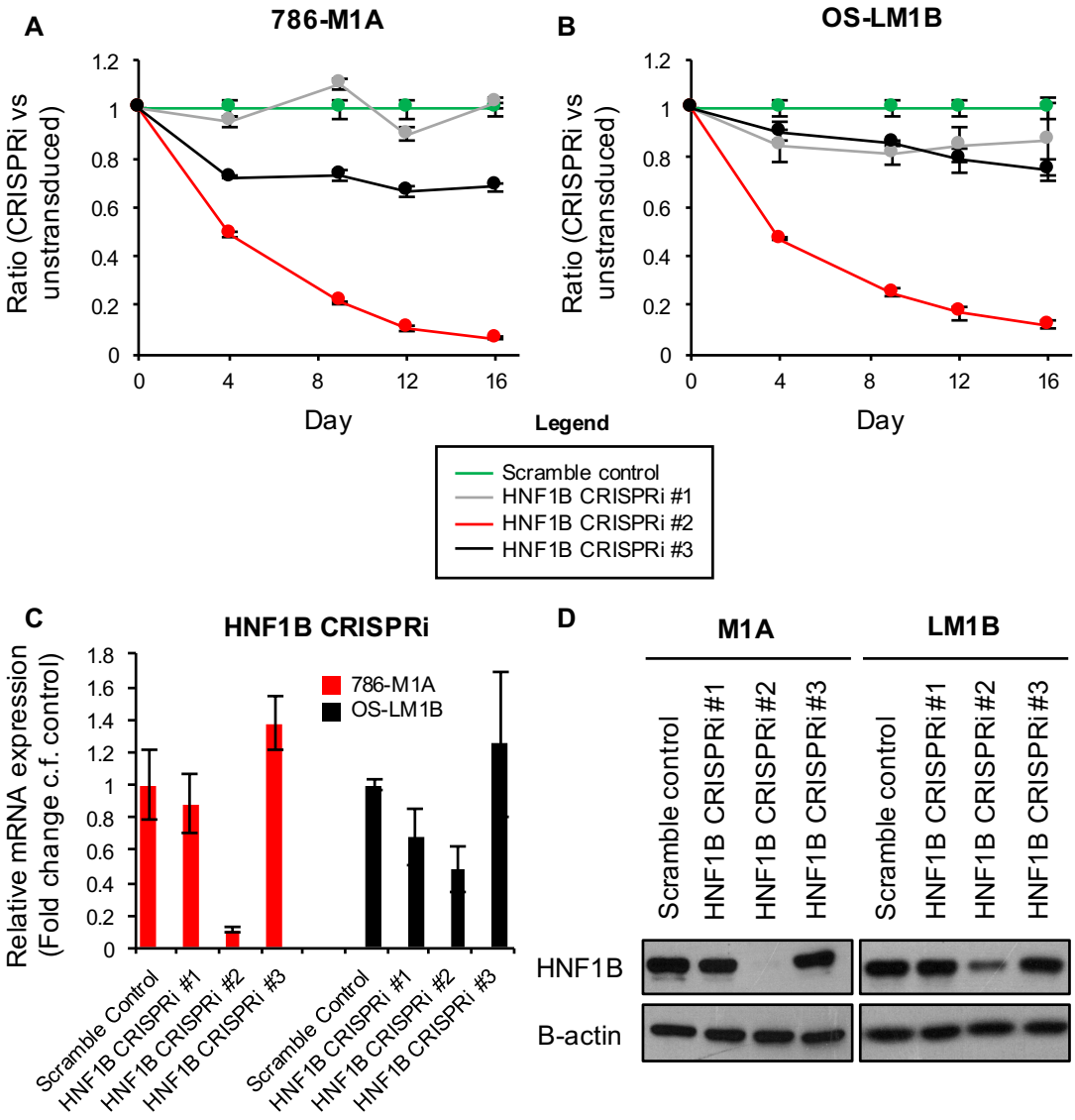


Fig 24: ***CRISPRi-mediated knockdown of HNF1B confers the same selective disadvantage, suggesting it is not due to /double-stranded break formation by Cas9.***

(A)-(B) Competition proliferation assay data showing the selective disadvantage conferred by HNF1B CRISPRi knockdown in (*A*) 786-M1A-dCas9-KRAB-mCherry cells and (*B*) OS-LM1B-dCas9-KRAB-mCherry cells. Untransduced parental BFP– cells (not transduced with sgRNAs) were mixed in a 50:50 ratio with BFP+ cells transduced with either the Scramble Control, HNF1B CRISPRi #1, HNF1B CRISPRi #2 or HNF1B CRISPRi #3 constructs. The percentages of BFP positive versus BFP negative cells were measured every 4 days by FACs, and used to calculate the BFP+/BFP– ratio over time.

(C) HNF1B qPCR data for the indicated 786-M1A-dCas9-KRAB-mCherry cell lines (*red*) and OS-LM1B-dCas9-KRAB-mCherry cell lines (*black*) used in the CRISPRi competition assays. HNF1B mRNA levels in each sample were normalised to the Scramble Control sample for each set of cell lines. The HNF1B CRISPRi #2 tandem sgRNA construct is the only one that successfully mediates HNF1B knockdown.

(D) Western blots for HNF1B in whole cell protein lysates from the indicated 786-M1A-dCas9-KRAB-mCherry cell lines (*left*) and OS-LM1B-dCas9-KRAB-mCherry cell lines (*right*) used in the CRISPRi competition assays. HNF1B is only knocked down in cells containing the HNF1B CRISPRi #2 construct.

protein levels similar to those seen in the Scramble Control cells (***Fig 24D***). This is seen for both the 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry sets of cell lines.

2.6) Investigating the phenotype for HNF1B CRISPR/Cas9 knockout in vivo

Taken together, the *in vitro* assays I performed established a clear phenotype for HNF1B loss. The next step was to conduct *in vivo* experiments to determine whether HNF1B is as important for tumour survival and tumour growth as it is for the survival and proliferation of ccRCC cell lines *in vitro*.

The parental 786-M1A and OS-LM1B cells, and thus each control or knockout cell line derived from them, contain a thymidine kinase-GFP-luciferase (TGL) transgene[122]. The luciferase allowed me to image tumour cells, and thus monitor tumour growth, *in vivo* by injecting luciferin and measuring the bioluminescent signal emitted.

To investigate whether HNF1B loss also produced a phenotype *in vivo*, I carried out fresh transductions of 786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cells with the Control-BFP, sg2-BFP and sg8-BFP constructs. I then used these cell lines to carry out a subcutaneous tumour xenograft assay in immunocompromised nude/athymic mice. This mouse strain lacks fur, which allows better subcutaneous tumour visibility and palpability. I prepared the 786-M1A-Cas9 and OS-LM1B-Cas9 control and knockout cells for injection by making up each cell line as a cell suspension in a mixture of sterile PBS and matrigel (50:50), at a concentration of $1x10^6$ cells/ml.

We had a total of 6 experimental groups (i.e. one group per cell line: 786-M1A-Cas9 Control, HNF1B sg2 and HNF1B sg8, and OS-LM1B-Cas9 Control, HNF1B sg2 and HNF1B sg8), with 5 mice per group. For each cell line, 100ul of cell suspension (a total of 100,000 cells) was injected subcutaneously into both the left and right flanks of each mouse, for a total of 10 injections per cell line.

Tumour growth was measured weekly by intravital bioluminescence imaging[4]: I administered D-luciferin by intraperitoneal injection (injecting a total of 100ul of 15mg/ml D-luciferin dissolved in 1X sterile PBS), and measured the resultant bioluminescent signal with the Xenogen IVIS. These bioluminescence measurements were then normalised to day 0 values and taken as a quantitative measure of tumour growth. Additionally, once the subcutaneous tumours became visible and palpable (in week 4 after injection), the bioluminescence measurements were supplemented with calliper measurements of tumour size (performed blind and by the same

individual on a weekly basis). These calliper measurements were used to calculate approximate tumour volume over time as an additional measure of tumour growth.

The experiment was allowed to continue until the largest Control tumours reached a mean diameter of 12mm. This was after 7 weeks of tumour growth. At this experimental endpoint, the animals were euthanized and I excised the subcutaneous tumours, excluding as much normal murine tissue as possible. I weighed the tumours to determine final tumour mass, and then fixed several tumours per experimental group in 10% formalin for immunohistochemistry.

The bioluminescence imaging data for both the 786-M1A-Cas9 (***Fig 25A***) and OS-LM1B-Cas9 (***Fig 25B***) subcutaneous tumour xenograft assays was consistent with the *in vitro* proliferation and competition assay data: control tumours grew rapidly, whereas HNF1B knockout tumours (with either sg2 or sg8) exhibited markedly impaired growth (***Fig 25A-B***). This difference between the control and knockout was especially striking in the OS-LM1B tumours, and indeed the *in vitro* phenotype was also more pronounced in OS-LM1B cells.

Similarly, the calliper measurements of tumour volume for both the 786-M1A-Cas9 (***Fig 25C***) and OS-LM1B-Cas9 (***Fig 25D***) subcutaneous tumour xenograft assays also showed a rapid increase in tumour size in the control tumours, but greatly reduced growth in the HNF1B knockout tumours, especially in the OS-LM1B system (***Fig 25C-D***).

Correspondingly, if we look at the final tumour weights, the knockout tumours were significantly smaller compared to control tumours at the end of the experiment, in both the 786-M1A-Cas9 (***Fig 25E***) and OS-LM1B-Cas9 (***Fig 25F***) subcutaneous tumour xenograft assays.

Collectively, these three sets of *in vivo* tumour growth data establish that the selective disadvantage conferred by HNF1B knockout *in vitro* is also observed *in vivo*, further underscoring the importance of HNF1B in ccRCC.

To confirm that the control and knockout cell lines injected for each *in vivo* experiment were indeed HNF1B wild-type and HNF1B-deficient respectively, I performed HNF1B Western blots with whole cell protein lysates from the cells used for the subcutaneous tumour xenograft assays. I took cell pellets for protein extraction from the freshly transduced 786-M1A-Cas9 and OS-LM1B-Cas9 Control-BFP, sg2-BFP and sg8-BFP cells, 4 days prior to subcutaneous injection, to determine the levels of HNF1B expression (or lack thereof) at a time point as close as possible to the point of tumour implantation.

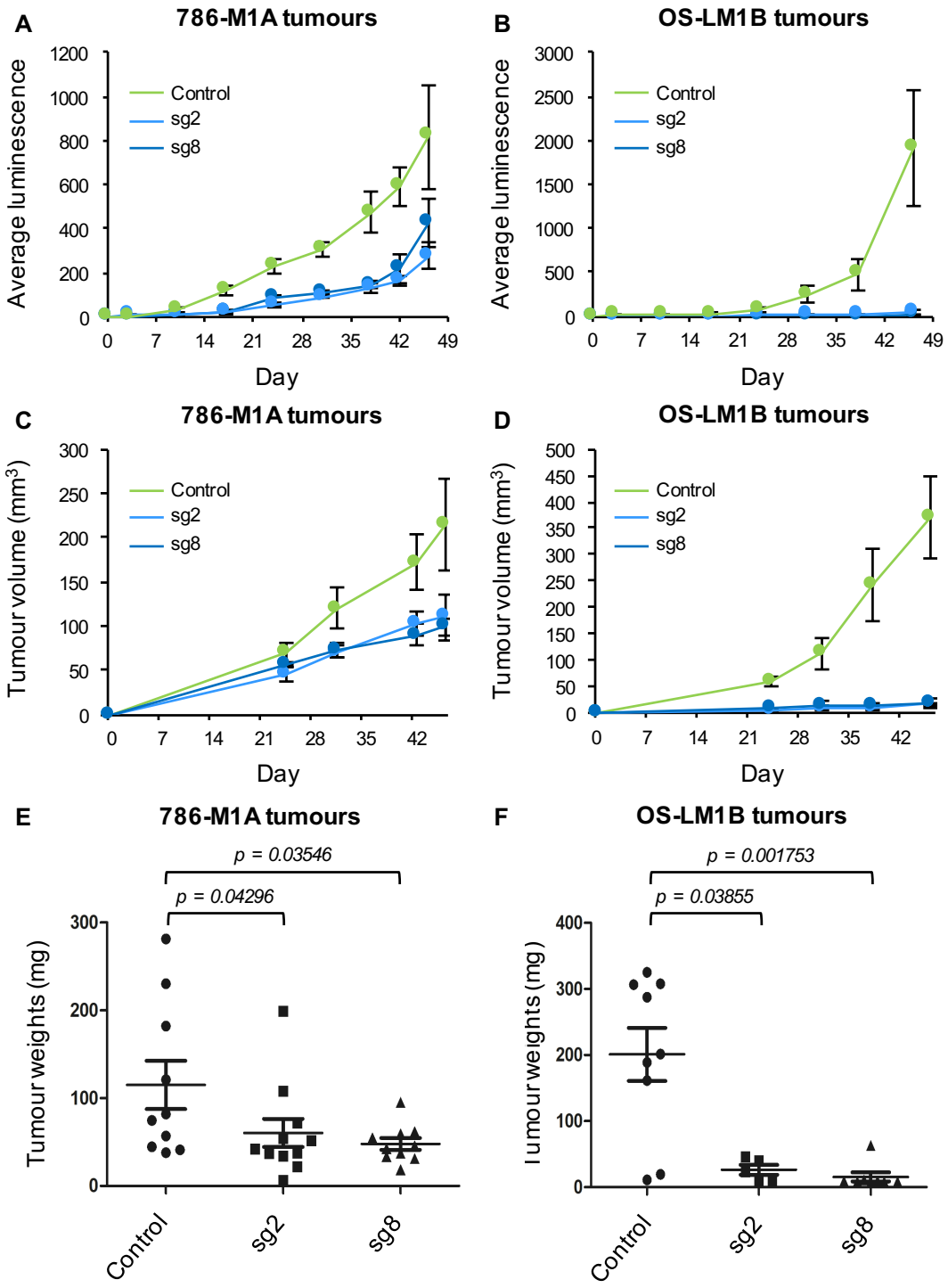


Fig 25: ***HNF1B knockout also impairs tumour initiation and growth in vivo.***
(A)-(B) Bioluminescence measurements of 786-M1A-Cas9 and OS-LM1B-Cas9 Control and HNF1B knockout (sg2 and sg8) tumour growth in each subcutaneous tumour xenograft assay.
(C)-(D) Calliper measurements of tumour volume for 786-M1A-Cas9 and OS-LM1B-Cas9 Control and HNF1B knockout (sg2 and sg8) tumours, once they became palpable on Day 24 of each subcutaneous tumour xenograft assay.
(E)-(F) Final tumour weights for the 786-M1A-Cas9 and OS-LM1B-Cas9 Control and HNF1B knockout (sg2 and sg8) tumours at the endpoint of each subcutaneous tumour xenograft assay.

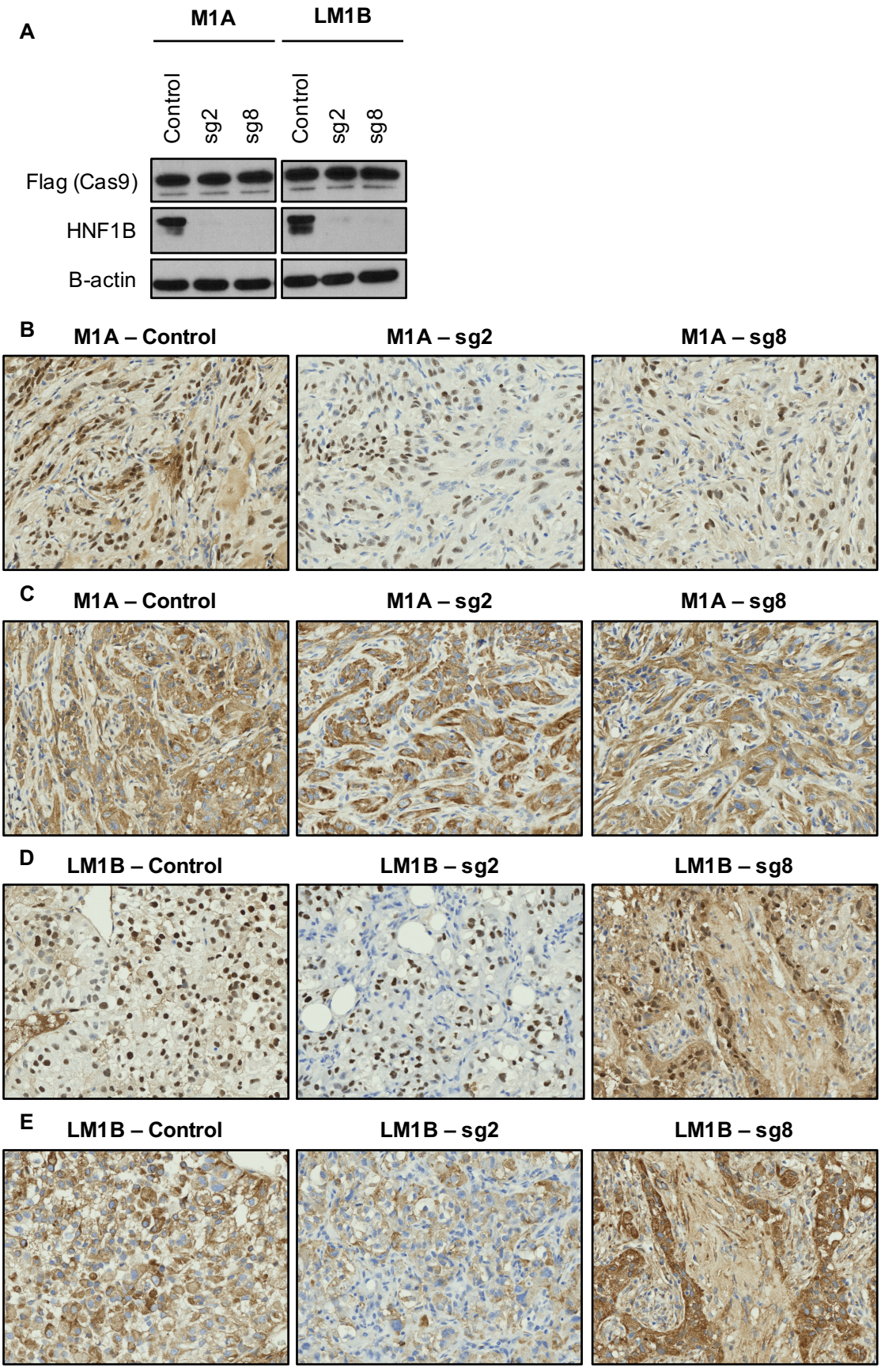


Fig 26: ***Characterisation of HNF1B expression in ccRCC cell lines prior to injection, and in the resultant tumours.*** P.T.O for figure legend

Fig 26: ***Characterisation of HNF1B expression in ccRCC cell lines prior to injection, and in the resultant tumours.***
(A) Western blots for HNF1B in whole cell protein lysates extracted 4 days prior to injection from the 786-M1A-Cas9 and OS-LM1B-Cas9 Control-BFP, sg2-BFP and sg8-BFP cell lines used for the subcutaneous tumour xenograft assay.

(B) HNF1B immunohistochemistry in tissue sections from 786-M1A-Cas9 Control (*left*), sg2 (*middle*), and sg8 (*right*) subcutaneous tumours from the xenograft assay.
(C) Vimentin immunohistochemistry in tissue sections from the same 786-M1A-Cas9 Control (*left*), sg2 (*middle*), and sg8 (*right*) subcutaneous tumours that were stained for HNF1B.

(D) HNF1B immunohistochemistry in tissue sections from OS-LM1B-Cas9 Control (*left*), sg2 (*middle*), and sg8 (*right*) subcutaneous tumours from the xenograft assay.
(E) Vimentin immunohistochemistry in tissue sections from the same OS-LM1B-Cas9 Control (*left*), sg2 (*middle*), and sg8 (*right*) subcutaneous tumours that were stained for HNF1B.

The HNF1B Western blot data verified that, in both sets of injected cell lines, HNF1B was strongly expressed in cells transduced with the non-targeting control sgRNA, and almost completely depleted in cells transduced with HNF1B sg2 or sg8, prior to the start of each *in vivo* experiment (***Fig 26A***).

Finally, I wanted to determine the levels of HNF1B expression in the control and knockout tumours at the end of each experiment. Firstly, I wanted to confirm whether the endpoint control tumours were still HNF1B-positive. Secondly, I wanted to ascertain whether the small tumours formed from knockout cells were entirely HNF1B-negative (and had thus become HNF1B-independent, and somehow circumvented the effects of HNF1B loss), or whether they were in fact entirely HNF1B-positive (and thus consisted of cells that had not undergone gene knockout), or whether these "knockout" tumours were comprised of both HNF1B-positive and HNF1B-negative escapers.

To address these questions, I used the excised 786-M1A-Cas9 and OS-LM1B-Cas9 control and HNF1B knockout (sg2 and sg8) tumours that I had fixed in 10% formalin at the end of each subcutaneous tumour xenograft assay. Each tumour was fixed in 10% formalin overnight, then washed with 1X PBS and transferred to 70% Ethanol (EtOH; *see also Materials and Methods*). The fixed tumours were then used to make paraffin-embedded tissue blocks, which were cut into sections and placed on slides for immunohistochemistry (IHC; all courtesy of the Addenbrooke's Hospital Tissue Bank service).

I then performed HNF1B IHC with tumour sections from 786-M1A-Cas9 and OS-LM1B-Cas9 Control and knockout (sg2 and sg8) tumours (***Fig 26B, D***). I used the Leica Bond machine to carry out all my IHC stainings, and used the Zeiss Axioscan to scan and image the slides afterwards. I used the HNF1B antibody from the Human Protein Atlas (HPA, supplied by Sigma) at an antibody concentration of 1:200, diluted in Bond buffer solution (*see also Materials and Methods*). Since HNF1B is a transcription factor, one would expect to see nuclear staining for HNF1B.

Control tumours exhibited stronger staining for HNF1B, with a greater frequency of HNF1B-positive nuclei that also stained more intensely (***Fig 26B, D; left image in each panel***), compared to the sg2 and sg8 knockout tumours (***Fig 26B, D; middle & right images in each panel***). Therefore, the endpoint control tumours were indeed still HNF1B-positive. However, the control tumours also contained some HNF1B-negative nuclei (***Fig 26B, D; left image in each***

panel), with varying proportions of these HNF1B-negative cells between different control tumours (*data not shown*), but always a minority of HNF1B-negative cells relative to the HNF1B-positive cells.

The sg2 and sg8 tumours also showed a mixture of HNF1B-positive and HNF1B-negative nuclei (***Fig 26B, D; middle & right images in each panel***). However, relative to the Control tumours, the HNF1B knockout tumours exhibited a greater frequency of HNF1B-negative nuclei, and correspondingly a lower frequency of HNF1B-positive nuclei that also stained less intensely with HNF1B antibody (***Fig 26B, D; middle & right images in each panel***).

The presence of both HNF1B-positive and HNF1B-negative nuclei in the endpoint sg2 and sg8 tumours could suggest that these tumours were comprised of both kinds of escapers outlined above. In other words, it is possible that the sg2 and sg8 tumours consisted of a mixture of HNF1B-negative ccRCC cells that had escaped the effects of the sgRNAs by becoming resistant to HNF1B depletion, and HNF1B-positive ccRCC cells that had avoided HNF1B knockout altogether (for instance by silencing Cas9 or sgRNA expression, or not forming frameshift-inducing indel mutations). However, it is also possible that the tumours comprised a mixture of ccRCC cells and murine cells from the surrounding mouse tissue, and that the HNF1B-negative nuclei thus corresponded to non-ccRCC murine cells instead. Indeed, this possibility might also explain the presence of HNF1B-negative nuclei in the Control tumours.

Therefore, I also stained tumour sections from the same tumours for human vimentin (***Fig 26C, E***). The cell lines I used for my experiments (and thus the cells in the subcutaneous tumours) are human ccRCC cell lines that have partially undergone the process of epithelial-to-mesenchymal transition (EMT), and thus all of them express vimentin. Staining for vimentin thus enabled me to confirm whether the tumours I excised were comprised of human ccRCC cells or murine cells (for instance, from the surrounding mouse stromal tissue).

Once again I used the Leica Bond machine to carry out the IHC stainings, and used the Zeiss Axioscan to scan and image the stained slides. I used the Vimentin antibody from Cell Signaling (supplied by NEB) at a 1:100 antibody dilution in Bond buffer (*see also Materials and Methods*). Since vimentin is a component of the cytoskeleton, one would expect to see cytoplasmic staining for vimentin.

The Control, sg2, and sg8 tumours all showed a mixture of vimentin-positive and vimentin-negative cells (***Fig 26C, E***). The presence of vimentin-negative cells suggests that the

subcutaneous tumours may have incorporated murine stromal cells into the tumour mass through various mechanisms as they grew. For example, the vimentin-negative cells may have belonged to blood vessels (since ccRCC tumours are usually highly vascular), lymphatics, fibroblasts, or even immune cells. Although nude athymic mice lack mature T cells, they still contain mature B cells, dendritic cells, macrophages and natural killer cells, and therefore might still have mounted some form of an immune response against the human cell line-derived subcutaneous tumours.

These potential identities of the vimentin-negative cells are certainly applicable to the Control tumours, which exhibited rapid growth *in vivo*. These possibilities are also applicable to the sg2 and sg8 tumours if they were indeed comprised of cells that had adapted to HNF1B loss or escaped HNF1B knockout. Such escapers would no longer experience a selective disadvantage from sgRNA expression and would thus have been able to grow, incorporating murine stromal cells to establish a murine-derived tumour microenvironment, to form the small tumours obtained at the end of the experiment.

In summary, it is likely that the HNF1B-negative cells in the Control tumours corresponded to non-ccRCC murine cells within the tumour mass. It is also possible that the HNF1B-negative cells in the sg2 and sg8 tumours comprised of both HNF1B knockout ccRCC cells that became HNF1B-independent and overcame the effects of HNF1B loss, and non-ccRCC murine cells that were incorporated into the tumours as escaper cell populations emerged and began to grow.

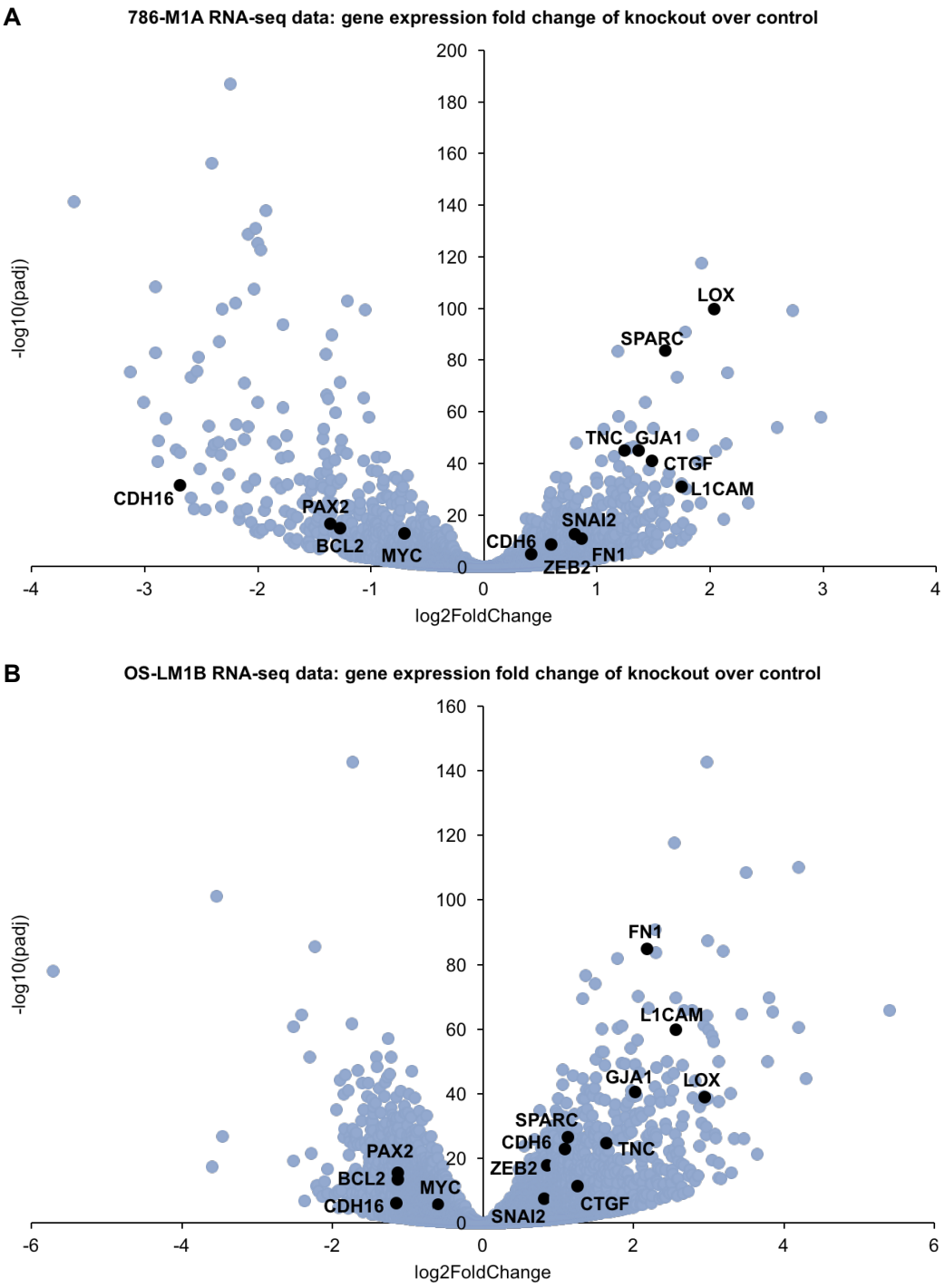


Fig 27: ***RNA-seq in HNF1B knockout versus control cells shows a downregulation of MYC and BCL2 upon HNF1B loss.***

(A)-(B) RNA-seq data showing gene expression changes in *(A)* 786-M1A-Cas9 knockout (sg2-BFP and sg8-BFP) cells relative to control (NTC18-BFP) cells, and *(B)* OS-LM1B-Cas9 knockout (sg2-BFP and sg8-BFP) cells relative to control (NTC18-BFP) cells. In both cell line systems, there was an upregulation of genes associated with epithelial-to-mesenchymal transition (EMT), a concomitant downregulation of the kidney epithelial marker CDH16, and a downregulation of MYC and BCL2 in HNF1B knockout cells relative to control cells.

(continued on next page)

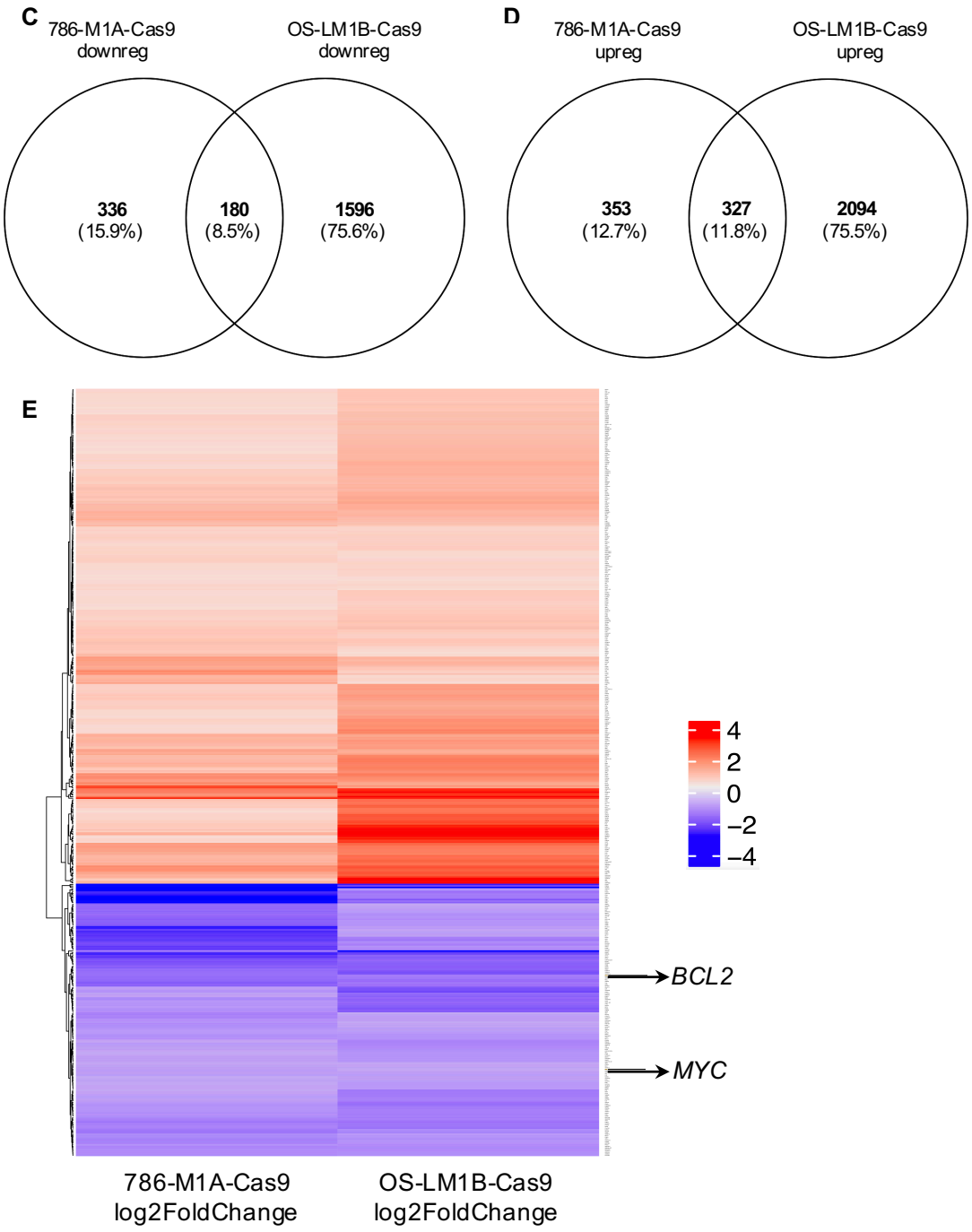
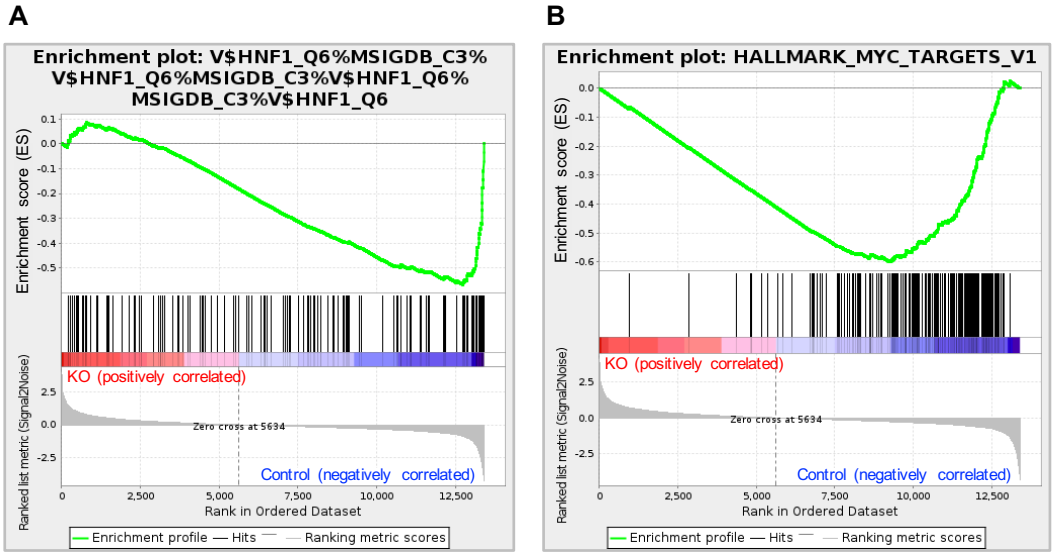


Fig 27 (continued): ***RNA-seq in HNF1B knockout versus control cells shows a downregulation of MYC and BCL2 upon HNF1B loss.***

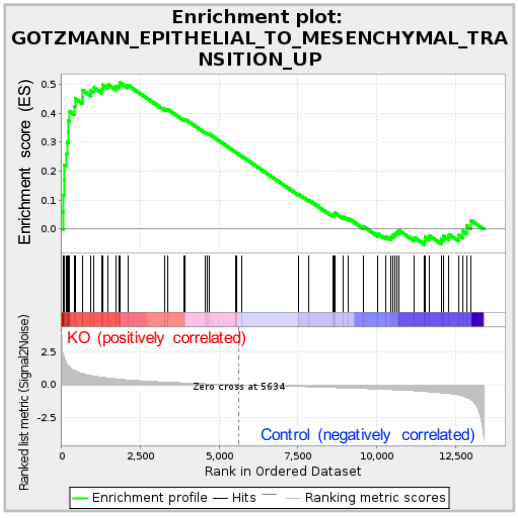
(C)-(D) Venn diagrams showing the number of *(C)* downregulated genes and *(D)* upregulated genes that are shared between the 786-M1A-Cas9 and OS-LM1B-Cas9 sets of RNA-seq data. These comparisons are based on those genes with an adjusted p-value (padj) ≤ 0.05, and a minimum 1.5 times fold change (i.e. $\log_2$ Fold Change ≥ 0.5849625 for upregulated genes, and $\log_2$ Fold Change ≤ -0.5849625 for downregulated genes). In total, there are 507 shared up-/down-regulated genes between 786-M1A-Cas9 and OS-LM1B-Cas9.

(E) Heatmap showing the 507 shared up-/down-regulated genes, and the magnitude of their expression changes (i.e. $\log_2$ Fold Change) in the 786-M1A-Cas9 and OS-LM1B-Cas9 RNA-seq datasets. The positions of MYC and BCL2 in the heatmap are indicated.



FDR q-value = 0.01

FDR q-value = 0.08889413



FDR q-value = 0.04122771

Fig 28: ***Gene set enrichment analysis (GSEA) of the RNA-seq data from the 786-M1A-Cas9 Control and HNF1B knockout (sg2 and sg8) cells.***

Representative gene set enrichment plots indicating significant signatures in the RNA-seq data for (A) an enrichment of HNF1B target genes in control versus knockout cells (indeed, one would expect HNF1B knockout cells to be relatively less enriched in HNF1B target gene expression), (B) an enrichment of MYC target genes in control versus knockout cells (consistent with the downregulation of MYC observed in the RNA-seq data in HNF1B knockout cells relative to controls), and (C) an enrichment of EMT-related genes in knockout versus control cells (as seen in the RNA-seq data). This gene set enrichment data is from the 786-M1A-Cas9 RNA-seq, but is representative of both the 786-M1A-Cas9 and OS-LM1B-Cas9 RNA-seq data.

CHAPTER 3: Characterising the phenotype of HNF1B loss and investigating the potential role of HNF1B in ccRCC

3.1) Identifying HNF1B target genes that may be involved in its role in ccRCC

After establishing the phenotype for HNF1B knockout *in vitro* and *in vivo*, the next step was to try to elucidate the mechanisms by which HNF1B supports ccRCC. I thus performed RNA-seq with 786-M1A-Cas9 and OS-LM1B-Cas9 knockout (sg2-BFP and sg8-BFP) and control (NTC18-BFP) cell lines to identify downstream transcriptional targets of HNF1B that may contribute to its critical role in ccRCC (***Fig 27A-B***).

I used the same batch of transduced cells as those used for the subcutaneous tumour xenograft assay. Therefore, these cell lines were already confirmed by Western blot to be HNF1B wild-type for the controls, and HNF1B-deficient for the knockouts (***Fig 26A***). I first expanded the 786-M1A-Cas9 and OS-LM1B-Cas9 Control, sg2, and sg8 cells to obtain four 10cm dishes each (in order to have 4 independent samples per cell line for RNA-seq) and extracted the RNA from each sample (see *Materials and Methods*). The RNA-seq libraries were then prepared in the lab with my RNA samples, and submitted for RNA-seq.

The sequencing reads from the RNA-seq were aligned to hg38 using RSEM[128] and bowtie2[129] with default settings. The alignment of the RNA-seq sequencing reads to hg38, as well as the differential expression (DEseq2) analysis and gene set enrichment analysis (GSEA) of the RNA-seq sequencing data, were all performed in the lab in collaboration with bioinformaticians from the Samarjiwa lab. For both the 786-M1A-Cas9 and OS-LM1B-Cas9 sets of cell lines, the data for sg2 and sg8 was combined and then compared to the control using DEseq2, to identify changes that were significant and consistent between both sgRNAs (***Fig 27A-B***). The gene expression changes were then compared between the 786-M1A-Cas9 and OS-LM1B-Cas9 datasets to discern which upregulated or downregulated genes were common to both cell lines (***Fig 27C-E***).

Since the OS-LM1B-Cas9 knockout cells experience a more extreme proliferative disadvantage relative to 786-M1A-Cas9 knockout cells (*see* ***Fig 14A-B*** *and* ***Fig 16A-B***), these cell lines took longer to expand to sufficient numbers. Consequently, the RNA-seq samples for the OS-LM1B-Cas9 cell lines were prepared 8 days after the RNA-seq samples for the 786-M1A-Cas9 cell lines (***Fig 30G***), and might thus reflect slightly later stage transcriptional changes

following HNF1B knockout. Nevertheless, the changes in expression of certain genes of interest were similar between both cell line systems (***Fig 27***).

In both cell line systems, there was an upregulation of genes associated with epithelial-to-mesenchymal transition (EMT) (***Fig 27A-B***). Consistent with this, there was also a downregulation of the kidney epithelial marker CDH16 (***Fig 27A-B***). HNF1B is a transcription factor that regulates and maintains an epithelial cell state in renal epithelial cells during kidney development[97–101,103,104]. Thus, one would expect to see the concomitant loss of epithelial cell markers upon knocking out HNF1B.

Interestingly, there was also a downregulation of MYC and BCL2 (***Fig 27A-B***), both or either of which may contribute to the selective disadvantage associated with HNF1B depletion. MYC promotes progression through the cell cycle, and is a key oncogene in a variety of different cancer types, where MYC overexpression enhances cell proliferation and tumour growth[130–132]. Thus, reduced MYC expression following HNF1B loss might contribute to the proliferative disadvantage seen in knockout ccRCC cell lines and tumours. Meanwhile, BCL2 is an anti-apoptotic protein. Thus, a decrease in BCL2 expression upon HNF1B loss could result in increased cell death, which might also contribute to the selective disadvantage observed with HNF1B knockout.

If we only consider those genes that are differentially expressed with an adjusted p-value (padj) ≤ 0.05 and a minimum 1.5 times fold change (i.e. a $\log_2$ Fold Change ≥ 0.5849625 for upregulated genes, and a $\log_2$ Fold Change ≤ -0.5849625 for downregulated genes), then we are left with a list of 4197 genes for OS-LM1B-Cas9, and 1196 genes for 786-M1A-Cas9.

If we look at the downregulated genes (1776 genes in OS-LM1B-Cas9; 516 genes in 786-M1A-Cas9) and upregulated genes (2421 genes in OS-LM1B-Cas9; 680 genes in 786-M1A-Cas9) separately, then we see that there is an overlap of 180 shared downregulated genes (***Fig 27C***) and 327 shared upregulated genes (***Fig 27D***) between the two cell line systems. If we calculate the total number of genes between the downregulated and upregulated overlaps, there is a total of 507 commonly and significantly differentially expressed genes, which all change in the "same direction" (i.e. upregulated or downregulated in *both* the 786-M1A-Cas9 and OS-LM1B-Cas9 RNA-seq datasets).

The heatmap in ***Figure 27E*** shows the 507 shared up-/down-regulated genes (represented as rows in the heatmap), and the magnitude of their expression changes (i.e. the $\log_2$ Fold

Change) in the 786-M1A-Cas9 and OS-LM1B-Cas9 RNA-seq datasets (represented as columns in the heatmap). *MYC* and *BCL2*, the two key genes of interest in the RNA-seq data, are indicated by arrows.

Gene set enrichment analysis (GSEA) of the RNA-seq data revealed a significant signature for an enrichment of HNF1B target genes in control versus knockout cells (***Fig 28A***). Indeed, one would expect HNF1B knockout cells to be relatively less enriched in HNF1B target gene expression. This analysis also revealed an enrichment of MYC target genes in control versus knockout cells (***Fig 28B***), which is consistent with the downregulation of MYC observed in the RNA-seq data in HNF1B knockout cells relative to controls. There was also a significant signature for an enrichment of EMT-related genes in knockout versus control cells (as seen in the RNA-seq data) (***Fig 28C***). These results were consistent between both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell line systems. Note that the GSEA plots shown in ***Figure 28*** are from the 786-M1A-Cas9 RNA-seq data, but are representative of both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell line systems.

To further investigate HNF1B transcriptional targets that may be involved in HNF1B's function in ccRCC, and in particular to identify genes that are direct targets of HNF1B, the lab performed HNF1B transcription factor ChIP-seq in the parental ccRCC cell lines 786-O and OS-RC2, and their more metastatic derivatives, 786-M1A and OS-LM1B respectively[4]. The HNF1B ChIP-seq experiments were performed by another member of the lab, as previously published[133] (*see also Materials and Methods*), using one replicate for each of these four cell lines. HNF1B ChIP-seq would reveal the genomic loci where HNF1B binds in ccRCC, which could identify genes that may be direct transcriptional targets of HNF1B.

The HNF1B ChIP-seq sequencing data was analysed in the lab as follows (*see also Materials & Methods*): the raw ChIP-seq sequencing reads were aligned to hg38 using bowtie2[129], producing sam files, which were converted into sorted bam files using samtools[134]. HNF1B peaks were then called using MACS2[135] with the option -q 5e-2. The corresponding input samples from the ChIP experiment for each cell line were used as controls.

Any overlap between the peaks called for each cell line was then determined by taking the peak summits +/-50bp (bedops --range 50 --merge), and identifying regions that had a peak in at least two samples (i.e. in at least two out of the four cell lines); this was done using the multiIntersectBed function in bedtools. Thus, the final list of called peaks for HNF1B contained

peaks that were shared between at least two of the four samples. The gene annotation for the peaks was carried out using Homer[136]. The sequencing reads were first extended to the modal length of 250bp, and duplicates were removed; genome coverage was then calculated for each region of interest. The total peak numbers for each cell line were as follows: 7132 peaks for 786-O, 15641 peaks for OS-RC2, 4022 peaks for 786-M1A, and 28551 peaks for OS-LM1B.

Once HNF1B peaks had been identified based on the Illumina ChIP-seq sequencing data, transcription factor motif analysis was also carried out in the lab to determine which, if any, of the genomic regions containing HNF1B binding peaks were also enriched for HNF1B binding motifs. First, the MEME suite module Meme[137] was used to carry out *de novo* motif discovery in the genomic sequences for the subset of HNF1B peaks that were shared by all four cell line samples. To perform this analysis, the summits +/- 50bp, from all the regions that had a peak in all four samples, were used to look for *de novo* motifs. The MEME suite module Tomtom[138] was then used to compare the motifs identified by the Meme analysis against known motifs. This analysis revealed a motif, 5'-GTTAAWYATTAAC-3', which was (i) most likely to be overrepresented within the sequences of interest, and (ii) very similar to the known HNF1B transcription factor motif published in the literature (5'-GTTAATNATTAAC-3')[103].

Therefore, the combination of the HNF1B ChIP-seq data analysis and the subsequent HNF1B transcription factor motif analysis identified genomic regions that seemed to be bound by HNF1B protein in multiple ccRCC cell lines based on the ChIP-seq data, and that were also enriched in HNF1B transcription factor motifs.

No HNF1B binding peaks were identified near any of the EMT genes that were upregulated in HNF1B knockout cells (*data not shown*). However, HNF1B motif-enriched HNF1B binding peaks were identified around *MYC* (***Fig 29A***), *BCL2* (***Fig 29B***), and *CDH16* (***Fig 29C***), three of the genes that were downregulated upon HNF1B loss (***Fig 27A-B***).

The *MYC* locus contained the highest number of HNF1B motif-enriched binding peaks (***Fig 29A***). HNF1B peaks were called both upstream and downstream of *MYC*, and these peaks were distanced much farther from the *MYC* gene itself compared to the peaks found in the other two loci (***Fig 29A***). The HNF1B motif-enriched binding peaks identified around *BCL2* were much closer to the gene sequence, and were found upstream, downstream, and within the *BCL2* sequence itself (***Fig 29B***). Meanwhile, in the case of *CDH16*, only one HNF1B motif-enriched binding peak was detected, adjacent to the *CDH16* sequence (***Fig 29C***).

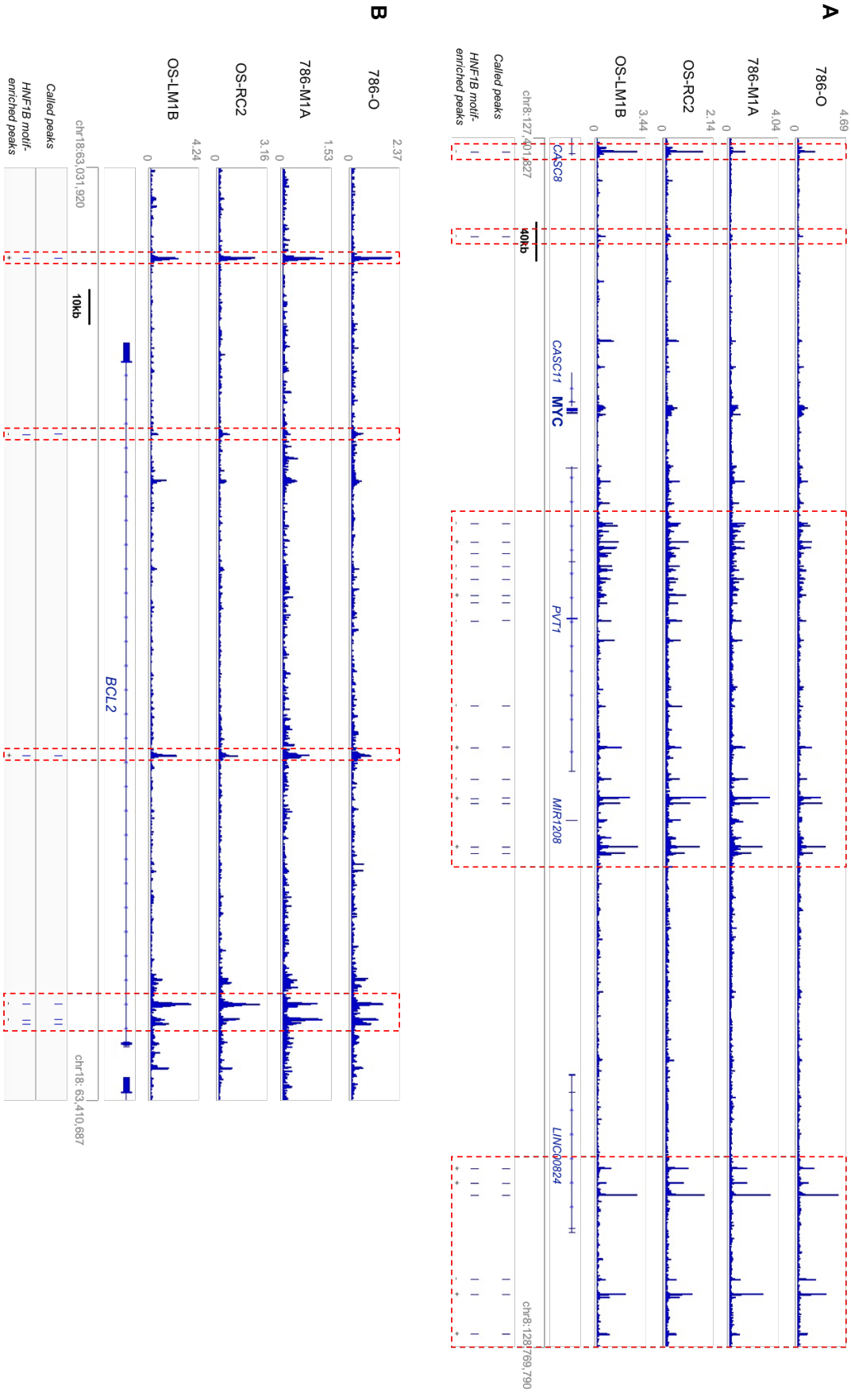


Fig 29: ***HNF1B transcription factor ChIP-seq in multiple ccRCC cell lines revealed HNF1B binding near the MYC, BCL2, and CDH16 loci.*** P.T.O for figure legend

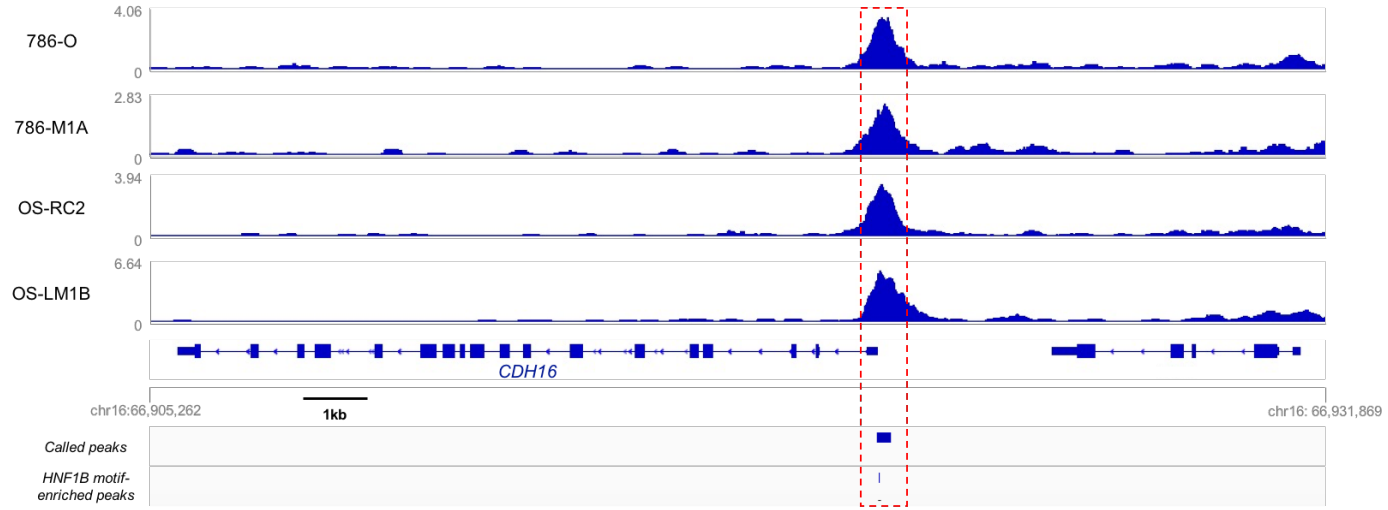


Fig 29: ***HNF1B transcription factor ChIP-seq in multiple ccRCC cell lines revealed HNF1B binding near the MYC, BCL2, and CDH16 loci.*** Data from HNF1B transcription factor ChIP-seq combined with HNF1B transcription factor motif analysis, performed in the four ccRCC cell lines indicated (786-O and OS-RC2, and their more metastatic derivatives 786-M1A and OS-LM1B respectively). HNF1B motif-enriched HNF1B binding peaks were identified near (*A*) *MYC*, (*B*) *BCL2*, and (*C*) *CDH16*, three of the genes that were downregulated upon HNF1B loss. Each HNF1B binding peak identified near these three genes (highlighted by the red dotted rectangles) was in a region that was both bound by HNF1B protein in all four cell lines based on the ChIP-seq data, and enriched in HNF1B transcription factor motifs.

Taking the data from the RNA-seq, HNF1B ChIP-seq, and HNF1B motif analysis together, MYC, BCL2 and CDH16 are likely to be *direct* transcriptional targets of HNF1B. Therefore, I focused on these 3 genes in subsequent experiments.

3.2) Validating the RNA-seq data by qPCR

To validate the results of the RNA-seq data, I used qPCR to confirm the downregulation of MYC, BCL2 and CDH16 in HNF1B knockout cells. For this qPCR validation, I used RNA samples that were extracted from cell pellets taken during the competition proliferation assays for the 786-M1A-Cas9 and OS-LM1B-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines.

I selected pellets from competition assay time points that roughly corresponded to cells that were in culture (and thus experiencing HNF1B knockout) for the same number of days post-infection (i.e. post-addition of virus) as the number of days post-infection for which the 786-M1A-Cas9 and OS-LM1B-Cas9 control and knockout cells were expanded before being harvested for RNA-seq (***Fig 30G***).

The 786-M1A-Cas9 control and knockout (control, sg2, and sg8) cells were prepared for RNA-seq at 10 days post-infection, whereas the OS-LM1B-Cas9 control and knockout cells were prepared for RNA-seq at 18 days post-infection. Thus, for the qPCR validation, I chose control, sg2, and sg8 RNA samples from competition assay Days 2 and 8 for the 786-M1A-Cas9 cell lines (corresponding to 12 and 18 days post-infection respectively), and Days 4 and 8 for the OS-LM1B-Cas9 cell lines (13 and 17 days post-infection) (***Fig 30G***).

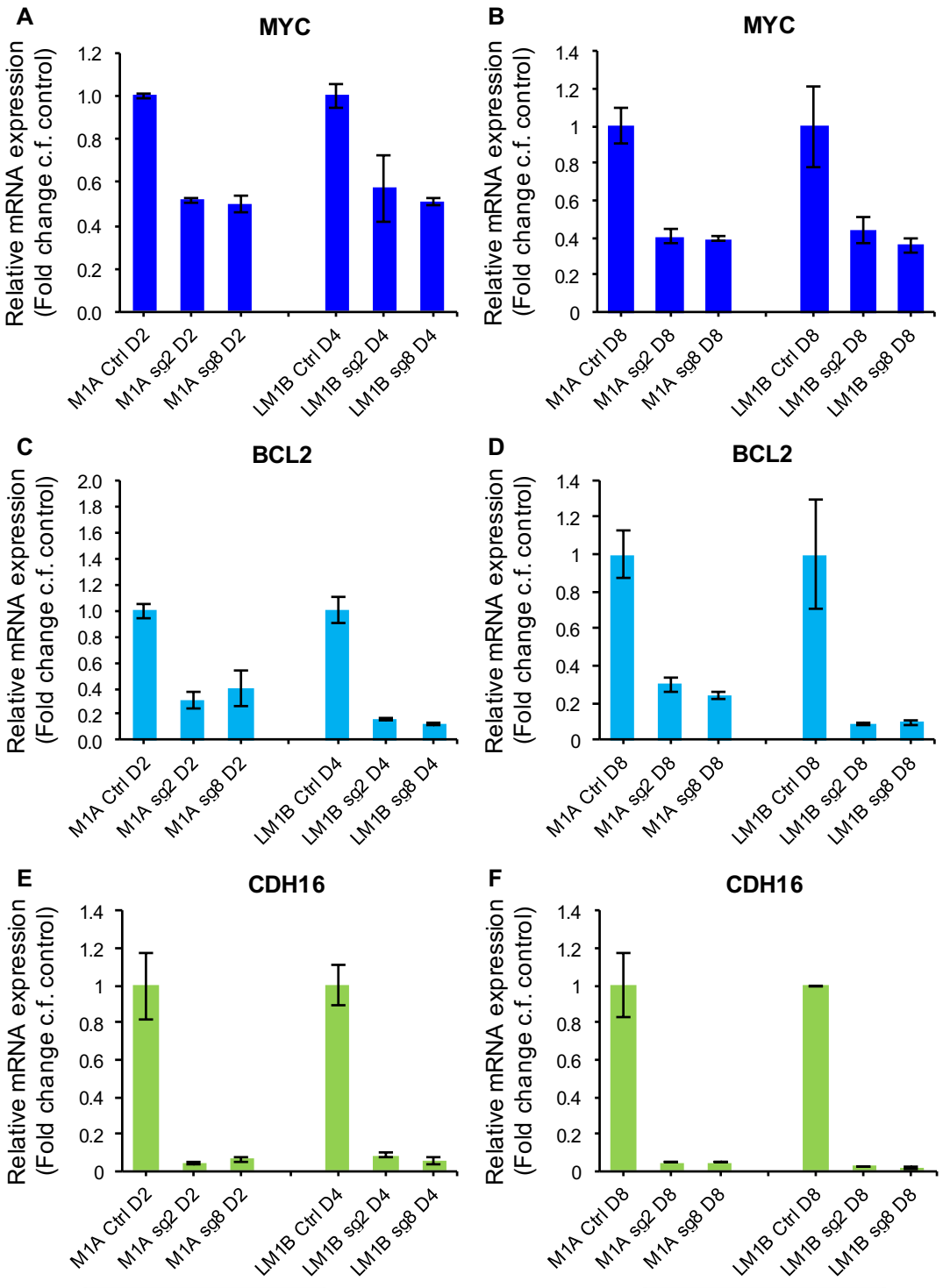
For the purposes of describing the qPCR validation data, I shall hereafter refer to the Day 2 and Day 4 time points (for the 786-M1A-Cas9 and OS-LM1B-Cas9 sets of cell lines respectively) as "early time points" and the Day 8 time points (for both sets of cell lines) as "late time points". Hence, the qPCR data for the early time points temporally matches the 786-M1A-Cas9 RNA-seq data, whereas the qPCR data for the late time points temporally matches the OS-LM1B-Cas9 RNA-seq data. However, as noted before, the gene expression changes observed by RNA-seq for genes of interest (in this case MYC, BCL2 and CDH16) were similar between the 786-M1A-Cas9 and OS-LM1B-Cas9 systems. Thus, *a priori*, one would expect to see similar trends in gene expression between the early and late time points in the results of the qPCR validation.

The average fold change in MYC expression in knockout versus control cells in the

RNA-seq data was 0.614 in the 786-M1A-Cas9 cells (a 38.6% decrease), and 0.662 (a 33.8% decrease) in the OS-LM1B-Cas9 Cells (***Fig 30G***). This is recapitulated in the MYC qPCR data for the early time points in both cell line systems (***Fig 30A***): 786-M1A-Cas9 knockout cells show an average fold change in MYC expression of 0.506 (i.e. a 49.4% decrease) relative to control cells, and OS-LM1B-Cas9 knockout cells show a similar average fold change of 0.540 (i.e. a 46% decrease) relative to controls. Meanwhile, in the late time points (***Fig 30B***), the 786-M1A-Cas9 and OS-LM1B-Cas9 knockout cells exhibit average MYC fold changes of 0.402 and 0.401 (i.e. 59.8% and 59.9% decreases in MYC mRNA levels) respectively, relative to controls. These values slightly exceed the changes seen in the RNA-seq data, but are still reasonably consistent.

In the case of BCL2, the RNA-seq data showed average fold changes of 0.414 (a 58.6% decrease) in 786-M1A-Cas9 knockout cells, and 0.456 (a 54.4% decrease) in OS-LM1B-Cas9 knockout cells (***Fig 30G***). The qPCR data for both early and late time points shows an even more dramatic downregulation for BCL2 in knockout cells, with average BCL2 fold changes of 0.357 (early) and 0.270 (late) in 786-M1A-Cas9 knockout cells (decreases of 64.3% and 73% respectively), and average fold changes of 0.142 (early) and 0.092 (late) in OS-LM1B-Cas9 knockout cells (decreases of 85.8% and 90.8% respectively) (***Fig 30C-D***). Despite the differences in the magnitude of downregulation observed, this data still validates that BCL2 is downregulated following HNF1B loss.

The expression data for CDH16 is consistent between both time points and both cell line systems in the qPCR, and is also consistent with (although somewhat exceeds) the fold changes seen in the RNA-seq data. The RNA-seq data indicated average fold decreases in CDH16 levels of 0.155 (an 84.5% reduction) in the 786-M1A-Cas9 knockout cells, and 0.451 (a 54.9% reduction) in the OS-LM1B-Cas9 knockout cells, relative to controls (***Fig 30G***). Meanwhile, the qPCR data showed average CDH16 fold changes of 0.054 (early) and 0.050 (late) in 786-M1A-Cas9 knockout cells (decreases of 94.6% and 95% respectively), and average fold changes of 0.071 (early) and 0.025 (late) in OS-LM1B-Cas9 knockout cells (decreases of 92.9% and 97.5% respectively) (***Fig 30E-F***). Thus, despite the different extents of downregulation observed between qPCR and RNA-seq, this data validates that there is a downregulation of CDH16 upon HNF1B loss.



G

786-M1A RNA-seq (10 days post-infection)	
Gene	*Fold change*
MYC	0.614
BCL2	0.414
CDH16	0.155

M1A
Day 2 = 12 days post infection
Day 8 = 18 days post infection
LM1B
Day 4 = 13 days post infection
Day 8 = 17 days post infection

OS-LM1B RNA-seq (18 days post-infection)	
Gene	*Fold change*
MYC	0.662
BCL2	0.456
CDH16	0.451

Fig 30: ***qPCR validation of the downregulation of MYC, BCL2, and CDH16 observed in the RNA-seq data.*** P.T.O for figure legend

Fig 30: ***qPCR validation of the downregulation of MYC, BCL2, and CDH16 observed in the RNA-seq data.***
(N.B. in the graphs above: M1A = 786-M1A-Cas9; LM1B = OS-LM1B-Cas9)

(A)-(B) MYC mRNA expression levels at the (*A*) early and (*B*) late time points, in 786-M1A-Cas9 and OS-LM1B-Cas9 control, sg2, and sg8 cell lines.

(C)-(D) BCL2 mRNA expression levels at the (*C*) early and (*D*) late time points, in 786-M1A-Cas9 and OS-LM1B-Cas9 control, sg2, and sg8 cell lines.

(E)-(F) CDH16 mRNA expression levels at the (*E*) early and (*F*) late time points, in 786-M1A-Cas9 and OS-LM1B-Cas9 control, sg2, and sg8 cell lines.

(G) Combined fold changes in gene expression (average of HNF1B sg2 and sg8 knockout cells compared to Control cells) for MYC, BCL2, and CDH16 from the RNA-seq data, in both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell lines. The number of days post-infection for which the cells were expanded before being harvested for RNA-seq are also indicated for each set of RNA-seq cell lines (*see the tables on the left and right*), and also for the cell lines used in the qPCR validation of the RNA-seq data (*see text in the middle*).

3.3) Investigating the mechanisms underlying the phenotype seen with HNF1B loss

In vitro, the selective disadvantage conferred by HNF1B depletion may be due to impaired progression through the cell cycle (and thus decreased cell proliferation), or increased cell death, or possibly a combination of both mechanisms. Therefore, the next step was to characterise the nature of the phenotype associated with HNF1B knockout. I thus carried out assays to check for increased apoptosis and defects in cell cycle progression in HNF1B knockout cells.

To measure apoptosis in each cell line, I stained control, knockout and rescue cells with Annexin V and Propidium Iodide, and analysed the cells by FACs to determine whether there was increased cell death in HNF1B knockout cells compared to controls, and if HNF1B restoration reversed this phenotype (*see also Materials and Methods*).

I also carried out cell cycle analysis experiments with Propidium Iodide staining to check for defects in cell cycle progression. I first treated control, knockout and rescue cells with 50ng/ml nocodazole for 24 hours to arrest the cells in the G2/M phase of the cell cycle, and thereby synchronise the cell cycles of the different cell lines. At the end of the 24hr treatment period, I took the cells off nocodazole to release the synchronised cells from mitotic arrest. I took samples (cell pellets) for each cell line every 2 hours for 16 hours, plus an eleventh sample at 24 hours post-release. Samples were also taken for untreated / unsynchronised (*UT unsync*) cells, and for the arrested / synchronised cells at the end of the nocodazole treatment period (*time point t0*), for each cell line. To achieve this time course, I seeded separate dishes per time point for each cell line. I fixed each of these cell pellet samples in 70% EtOH and froze them at -20'C. I then stained these fixed and frozen time course samples with Propidium Iodide, and analysed the cell cycle distribution at each time point by FACs to see (i) if I could observe any defects in cell cycle progression during the release period in HNF1B knockout cells relative to control cells, and (ii) if any such defect(s) were rescued in cells expressing the exogenous HNF1B transgene (*see also Materials and Methods*).

For the purposes of these apoptosis and cell cycle experiments, I thawed frozen stocks of the same 786-M1A-Cas9 control, knockout, and rescue cell lines that I characterised in the *in vitro* proliferation and competition assays (***Figs 14A, 18A, 18C, and 23A-B***), namely the 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cells (***see Figs 14A, 18A, and 18C***), and the

786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells (***see Fig 23A-B***).

For the Annexin V/Propidium Iodide (PI) staining, I used the eBioscience Annexin V Apoptosis Detection Kit APC from Thermo Fisher Scientific, with APC-conjugated Annexin V, which thus has an emission maximum of 660nm. Meanwhile, PI has an emission maximum of 617nm (once bound to DNA), and thus also emits in the red/infra-red region of the spectrum. However, the mCherry fluorescent protein has an emission maximum of 610nm. Therefore, there is considerable spectral overlap between PI and mCherry, as their emission maxima are very similar. In light of this, I had to use the BFP versions of each cell line (as listed above) for both the apoptosis and cell cycle experiments. For similar reasons, I was unable to perform these assays with any of the 786-M1A-dCas9-KRAB-mCherry CRISPRi cell lines used for the HNF1B CRISPRi competition assay (***Fig 24A***).

3.4) Potential mechanisms underlying the phenotype seen with HNF1B loss: cell cycle defects

The cell cycle assays with the 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines did not show much difference between the cell cycle distributions of untreated, unsynchronised Control, sg2 and sg8 cells (***Fig 31, UT unsync***). The control and knockout cell lines all showed normal cell cycle distributions: a prominent G0/G1 peak (the first peak, at approximately 50K fluorescence intensity) and a distinct, albeit smaller, G2/M peak (the second peak, at approximately 100K fluorescence intensity), with the cells in S phase lying between the two peaks.

At time point t0, immediately after 24 hours of nocodazole treatment, all three cell lines were arrested in the G2/M phase of the cell cycle (and were thus synchronised in their cell cycles), evident from the disappearance of the first (G0/G1) peak and the prominence of the second (G2/M) peak in the cell cycle distributions of the Control, sg2, and sg8 cells (***Fig 31, t0***).

Two hours after nocodazole release (time point t2), all three cell lines exited G2/M and progressed through the cell cycle into G0/G1, indicated by the decrease in size of the G2/M peak, and concomitant appearance of a prominent G0/G1 peak (***Fig 31, t2***). From time points t2 through to t8 (i.e. between 2 and 8 hours after release from mitotic arrest), the three cell lines exhibited relatively similar double-peaked cell cycle distributions, with no notable differences between knockout and control cells (***Fig 31, t2-t8***).

However, by time point t12 (i.e. at 12 hours after release), the sg2 and sg8 knockout cells seemed to be delayed in their cell cycle progression relative to the Control cells. At t12, while the majority of the Control cells appeared to be in S phase, both knockout cell lines still seemed to be mainly in G1, with a noticeable shoulder on the right hand slope of their G0/G1 peaks (***Fig 31, t12***). This could indicate a defect in cell cycle progression in HNF1B knockout cells relative to HNF1B wild-type controls, at the point of transitioning from G1 to S phase.

This impaired cell cycle progression in the knockout cells became even more obvious at t14, when the Control cells appeared to be entirely in S phase (***Fig 31, t14***). By contrast, in both the sg2 and sg8 cell lines there was still a portion of the cell population that seemed to be in G1, and the shoulder on the right hand side of the G0/G1 peak was even more pronounced, suggesting that the knockout cells were slower to enter S phase (***Fig 31, t14***). These differences in cell cycle distribution thus further indicated an impaired G1/S phase transition in HNF1B knockout cells relative to HNF1B wild-type controls.

This marked difference in cell cycle between control and knockout cells was also visible at the t16 time point. At t16, the majority of the sg2 and sg8 cells seemed to have progressed to S phase (***Fig 31, t16***). However, the Control cells were still ahead in their cell cycle progression at this time point, as the majority appeared to be entering the G2/M phase of the cell cycle (***Fig 31, t16***).

By time point t24 (24 hours after nocodazole release), the cell cycle distributions of all three cell lines returned to normal, with no obvious differences between control and HNF1B knockout cells (***Fig 31, t24***).

The cell cycle assays with the 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells supported similar conclusions regarding a potential cell cycle defect at the G1/S transition in HNF1B knockout cells (***Fig 32***). Note that in ***Figure 32***, cells containing empty vector are labelled "+EV", while those containing the HNF1B rescue construct are labelled "+rescue".

The cell cycle distributions of the three cell lines were again fairly similar when unsynchronised (***Fig 32, UT unsync***). After 24 hours of nocodazole treatment (t0), the majority of each cell population was in the G2/M phase of the cell cycle, indicating that all three cell lines underwent mitotic arrest (***Fig 32, t0***). However, the mitotic arrest was not as complete in this assay, as all three cell lines still retained a small G0/G1 peak.

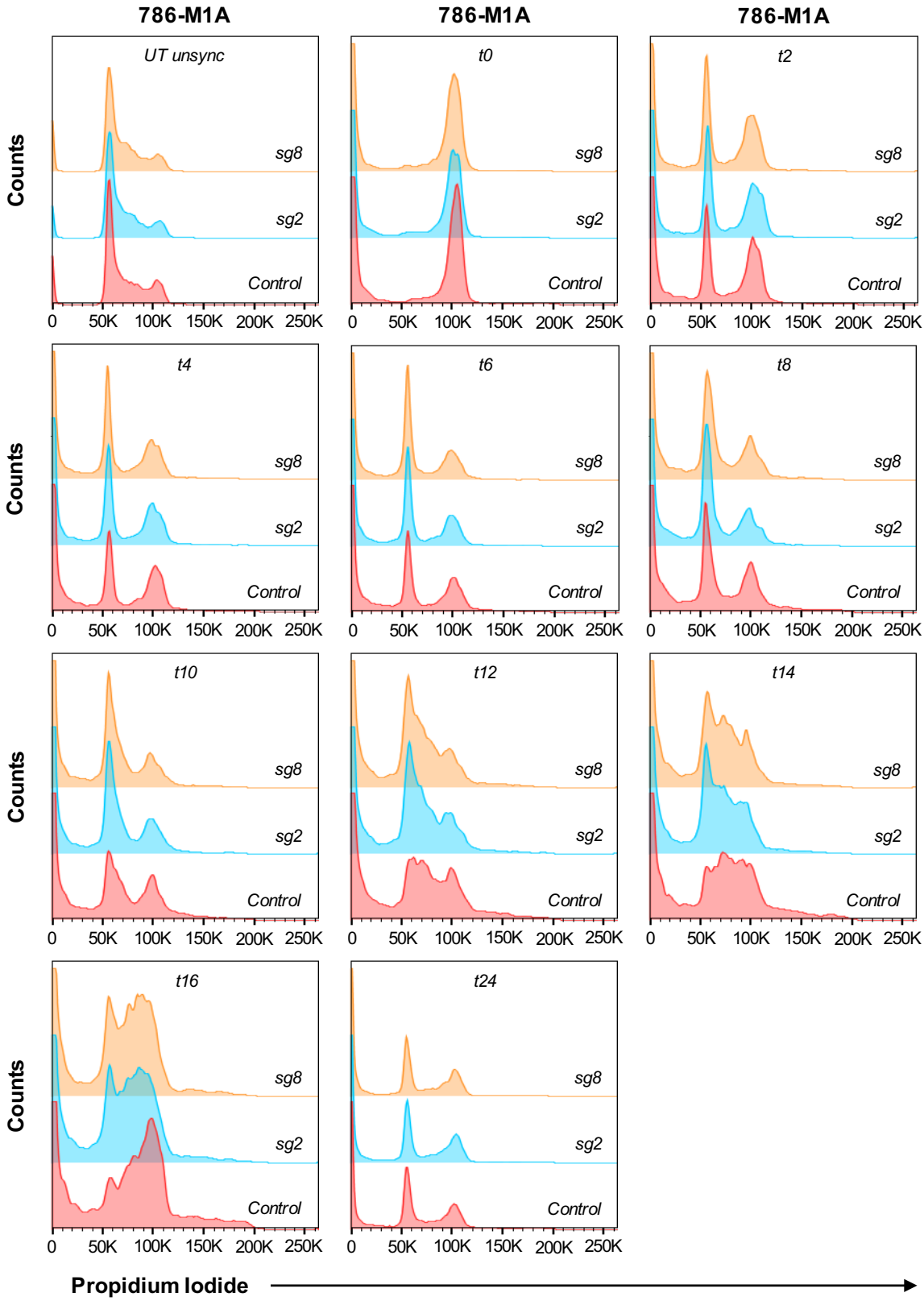


Fig 31: ***HNF1B knockout cells seem to exhibit delayed cell cycle progression relative to HNF1B wild-type controls, at the transition from G1 to S phase.*** FACs data for the cell cycle analysis of 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines, obtained via propidium iodide staining at the indicated time points after releasing the cells from mitotic arrest.

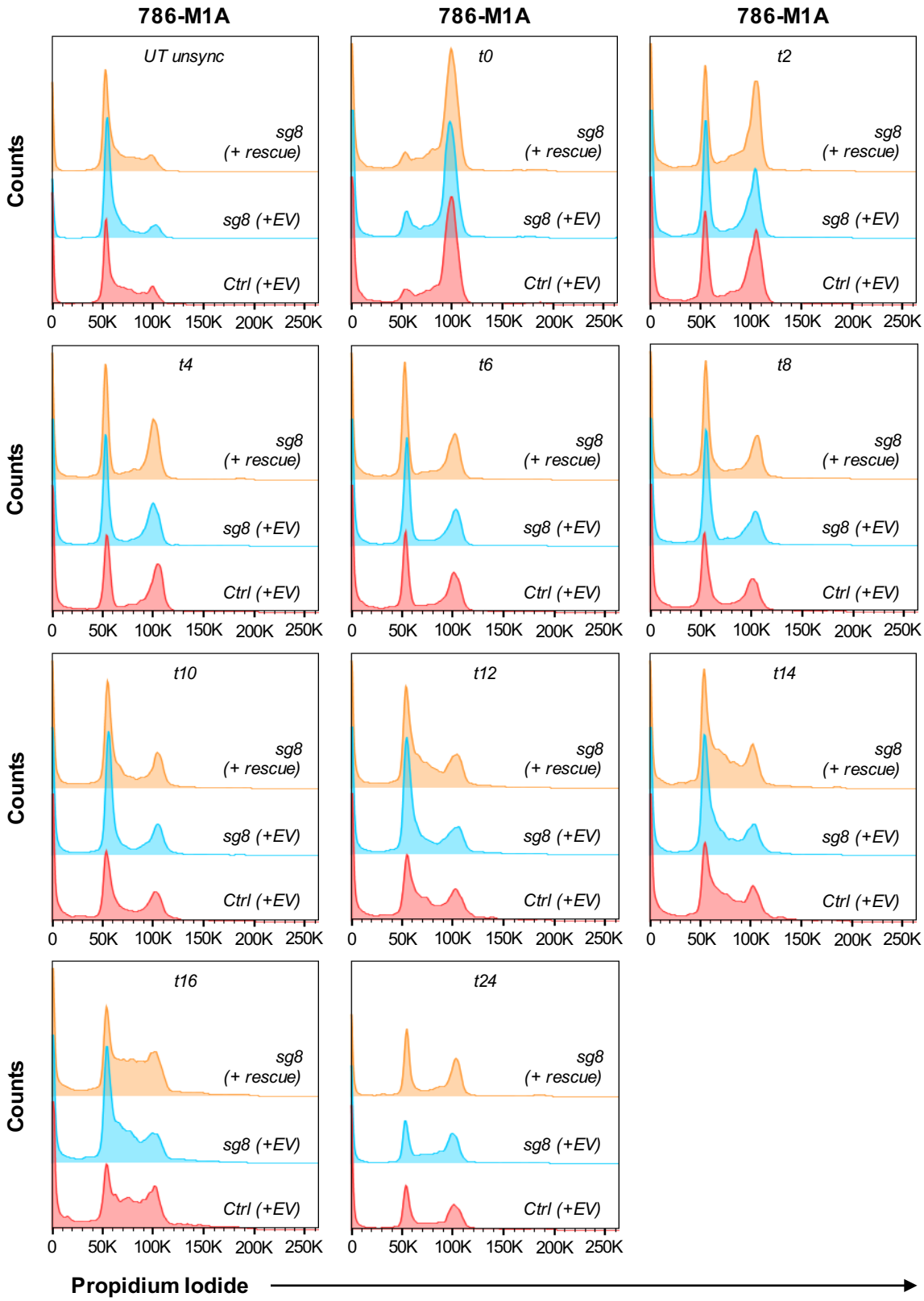


Fig 32: ***The delayed transition from G1 to S phase in HNF1B knockout cells is reversed by HNF1B restoration, via the expression of exogenous HNF1B.*** FACs data for the cell cycle analysis of 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cell lines, obtained via propidium iodide staining at the indicated time points after releasing the cells from mitotic arrest. "EV" indicates "empty vector"

Two hours after the removal of nocodazole (t2), all three cell lines were successfully released from G2/M and regained a prominent G0/G1 peak (***Fig 32, t2***). Subsequently, from time points t2 to t8, there were no striking differences between control, knockout, and rescue cells (***Fig 32, t2-8***).

However, similar to the other cell cycle assay comparing Control, sg2 and sg8 cells, differences in the cell cycle distributions of the HNF1B-depleted versus HNF1B wild-type cells started to emerge at t12. From t12 to t14, there appeared to be an increasing proportion of Ctrl (+EV) and sg8 (+rescue) cells exiting G1 and entering S phase (***Fig 32, t12-14***). This was suggested by the appearance and expansion of a shoulder on the right hand side of the G0/G1 peaks for both the control and rescue cell lines, and the concomitant expansion of the S phase region (lying between the G0/G1 and G2/M peaks) in the cell cycle distribution of each cell line (***Fig 32, t12-14***). By contrast, the sg8 (+EV) cells showed barely any transition out of G1 into S phase during this time period: the majority of knockout cells seemed to remain in G1, indicated by the prominent, undistorted G0/G1 peak across both time points (***Fig 32, t12-14***). Altogether, the cell cycle data from the t12 to t14 time points suggested that HNF1B knockout cells were experiencing a defect in exiting G1 and entering S phase relative to Control cells, and that this defect was reversed in the rescue cells.

The apparent cell cycle defect in HNF1B-depleted cells at the G1/S phase transition was also obvious at the t16 time point, when even more control and rescue cells appeared to have progressed into S phase (***Fig 32, t16***). Indeed, the region of the cell cycle distribution between the G0/G1 and G2/M peaks in both cell lines was distinctly raised and larger in size, which indicated an increase in the number of events detected at that intensity of PI fluorescence (***Fig 32, t16***). By contrast, the knockout cells seemed to be just starting to enter S phase at this late time point: the cell cycle distribution for the knockout cells had only just started to exhibit a discernible shoulder on the right hand slope of the G0/G1 peak (***Fig 32, t16***). This suggested that the knockout cells had only progressed to a stage in the cell cycle that was similar to the control and rescue cell cycles at t12 and t14 (***Fig 32, t12-16***). This data further supports the conclusion that HNF1B knockout cells have an impaired ability to exit G1 and enter S phase, which is recovered in the presence of exogenous HNF1B.

Finally, at t24 (24 hours after nocodazole release), all three cell lines seemed to have progressed through S phase into G2/M, and showed a more normal double-peaked cell cycle distribution (***Fig 32, t24***).

In summary, the data from both cell cycle assays suggested that HNF1B knockout cells exhibit delayed cell cycle progression relative to HNF1B wild-type cells, at the transition from G1 to S phase. Furthermore, as seen in the second cell cycle experiment, this delay in cell cycle progression is reversed by HNF1B restoration, via the expression of exogenous HNF1B.

3.5) Potential mechanisms underlying the phenotype seen with HNF1B loss: cell death

In the apoptosis experiments, Annexin V and Propidium Iodide staining distinguishes 3 cell populations: live cells, early apoptotic cells, and late apoptotic / necrotic (or already dead) cells.

Annexin V binds to phosphatidylserine (PS). PS is mainly located in the inner leaflet of the cell membrane (and is therefore inaccessible) in healthy viable cells, but becomes exposed on the outer leaflet of the cell membrane during the initial stages of apoptosis. Annexin V thus distinguishes between non-apoptotic and early apoptotic cells. Propidium Iodide (PI) meanwhile is a viability dye that binds to DNA, but is excluded from cells with an intact cell membrane, and is used to detect dead cells. PI staining thus helps to distinguish between live (i.e. viable, PI-negative) and dead (i.e. non-viable, PI-positive) cells, and also to differentiate between early- and late-stage apoptosis (and/or necrosis).

Live, healthy cells therefore stain as Annexin V-negative / PI-negative (Annexin V–/PI–). Cells undergoing early apoptosis still have an intact plasma membrane, and thus stain as Annexin V-positive / PI-negative (Annexin V+/PI–). However, cells undergoing late apoptosis or necrosis, or cells that are already dead, lose their plasma membrane integrity and therefore stain as Annexin V-positive / PI-positive (Annexin V+/PI+).

In the Annexin V / PI stainings performed with 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cells (***Fig 33A-C***), the majority (69.5%) of the Control cells were Annexin V–/PI– (i.e. viable and non-apoptotic), with 16.8% Annexin V+/PI+ (i.e. late apoptotic / necrotic / dead) cells (***Fig 33A***). By contrast, the cells expressing sg2 (***Fig 33B***) and sg8 (***Fig 33C***) had much smaller live cell populations (with 39.3% and 44.1% Annexin V–/PI– cells respectively) and showed increased percentages of late apoptotic / necrotic / dead cells (with 41.8% and 40.3%

Annexin V+/PI+ cells respectively), more than double the percentage seen in the Control. Thus, HNF1B loss does seem to result in increased levels of cell death relative to HNF1B wild-type controls.

The percentages of Annexin V+/PI– (i.e. early apoptotic) cells remained similar between the three cell lines, with 6.34%, 6.36% and 5.48% in the Control, sg2, and sg8 cells respectively. Thus, the main difference in the extent of cell death between control and HNF1B knockout cells appears to be in the proportion of late apoptotic / necrotic / dead cells.

Similar trends were observed in the Annexin V / PI stainings performed with 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells (***Fig 33D-F***). The majority (83.6%) of the Control (+ EV) cells were Annexin V–/PI– (i.e. viable and non-apoptotic), with only a small percentage (8.48%) of Annexin V+/PI+ (i.e. late apoptotic / necrotic / dead) cells (***Fig 33D***). By contrast, and consistent with the previous apoptosis experiment, the sg8 (+ EV) cell line exhibited a smaller live cell population (with 69.4% Annexin V–/PI– cells) and an increased percentage of late apoptotic / necrotic / dead cells (with 17.9% Annexin V+/PI+ cells), again more than double the percentage seen in the Control (***Fig 33E***). However, the sg8 (+ rescue) cells closely resembled the Control (+ EV) cells, with most of the cells (86.8%) staining as Annexin V–/PI– (i.e. viable and non-apoptotic), and a similarly small percentage (7.15%) of cells staining as Annexin V+/PI+ (i.e. late apoptotic / necrotic / dead) (***Fig 33F***). Once again, there was minimal change in the percentages of Annexin V+/PI– (i.e. early apoptotic) cells between the three cell lines, with 4.97%, 4.96% and 3.71% in the Control (+ EV), sg8 (+ EV), and sg8 (+ rescue) cells respectively.

Therefore, this data also suggests that HNF1B loss results in increased levels of cell death relative to HNF1B wild-type controls, and that this difference is mainly in the proportion of late apoptotic / necrotic / dead cells. Moreover, the increased level of cell death seen in HNF1B knockout cells appears to be rescued by HNF1B restoration, suggesting that the increased cell death is linked to HNF1B loss.

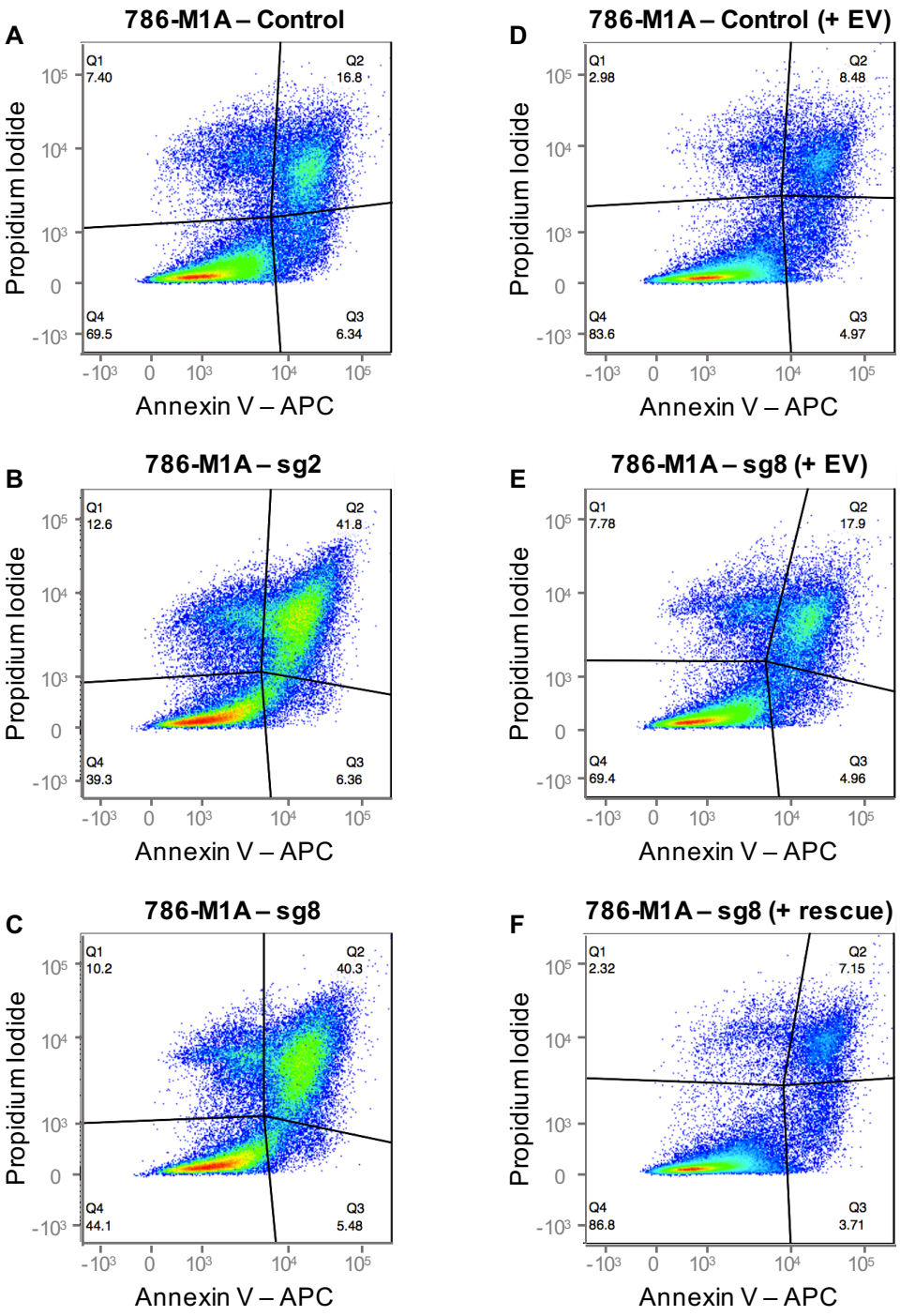


Fig 33: ***HNF1B knockout seems to be associated with increased cell death, and this phenotype is reversed by HNF1B restoration***

(A)-(C) FACs data from Annexin V / Propidium Iodide staining performed with 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cells to measure apoptosis levels.

(D)-(F) FACs data from Annexin V / Propidium Iodide staining performed with 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells to measure apoptosis levels.

Q2: Late apoptotic, necrotic, or already dead cells. Q3: Early apoptotic cells. Q4: Live, viable cells.

3.6) Investigating the contribution of MYC and BCL2 to the phenotype seen with HNF1B loss

Altogether, my cell cycle and apoptosis assays suggested that the selective disadvantage associated with HNF1B loss may be partly due to increased cell death, in addition to impaired cell cycle progression and decreased ccRCC cell proliferation. Therefore, HNF1B seems to play a role in promoting cell cycle progression and in preventing cell death in ccRCC.

Combining the cell cycle and apoptosis data with the RNA-seq in HNF1B knockout versus control cells, the HNF1B ChIP-seq and the HNF1B transcription factor motif analysis, my results highlighted the possibility that the decreased expression of the cell cycle protein MYC, and/or the decreased expression of the anti-apoptotic protein BCL2, contributed to the selective disadvantage associated with HNF1B depletion.

Indeed, MYC and BCL2 are both downregulated in HNF1B knockout cells based on my RNA-seq data. They also appear to be direct transcriptional targets of HNF1B, based on the HNF1B ChIP-seq and transcription factor motif analysis data. Moreover, MYC and BCL2 are involved in the cell cycle and apoptosis. Therefore, the importance of HNF1B in ccRCC may ultimately be linked to its ability to regulate MYC and BCL2 expression. Therefore, I decided to investigate MYC and BCL2 in subsequent experiments.

To investigate the relative contributions of MYC and BCL2 to the selective disadvantage associated with HNF1B depletion, I decided to test whether knockdown of MYC and BCL2 individually (or in combination) would phenocopy the effects of HNF1B knockdown.

It is possible that the depletion of either target gene alone would produce an equally strong phenotype as that conferred by HNF1B loss. Alternatively, each gene may only partially contribute to the selective disadvantage imposed by HNF1B depletion. HNF1B likely performs multiple functions in ccRCC, and indeed multiple genes are up- or down-regulated in response to HNF1B knockout, any or all of which may provide some contribution to the role of HNF1B and thus to the loss of cell fitness in its absence. If this is the case, individual knockdown of MYC or BCL2 would result in only a partial phenotype compared to that seen with HNF1B knockdown. Therefore, I tested the effects of knocking down MYC and BCL2 in combination as well as individually.

I designed tandem CRISPRi sgRNA constructs targeting MYC and BCL2 (two constructs each). As with the constructs used for the HNF1B CRISPRi (HNF1Bi) experiments (***Fig 24***), each of the MYC CRISPRi (MYCi) and BCL2 CRISPRi (BCL2i) tandem constructs comprised

two different sgRNAs targeting the gene in question (again specifically designed for CRISPRi applications), with each sgRNA flanked by its own U6 promoter and sgRNA scaffold, in a BFP fluorescent backbone. Thus, as before, the structure of each tandem construct was as follows: U6 promoter—CRISPRi sgRNA—sgRNA scaffold—U6 promoter—CRISPRi sgRNA—sgRNA scaffold. The CRISPRi sgRNA sequences were obtained from the Broad Institute online sgRNA design tool (https://portals.broadinstitute.org/gpp/public/).

In total, I designed two MYC-targeting constructs, MYCi(1) and MYCi(2) (containing different pairs of sgRNAs), and two BCL2-targeting constructs, BCL2i(3) and BCL2i(4) (also containing independent pairs of sgRNAs). As a control, I used the same "Scramble Control" tandem construct (containing two non-targeting "scramble" CRISPRi sgRNAs, also with a BFP fluorescent marker).

As with the HNF1Bi experiment, I used 786-M1A and OS-LM1B ccRCC cell lines stably expressing the dCas9-KRAB-mCherry transgene, which inhibits target gene transcription and thereby mediates CRISPRi knockdown. The BFP served as a marker of tandem sgRNA expression, while the mCherry served as a marker of dCas9-KRAB expression.

First, I tested the knockdown efficiency of each of the MYC and BCL2 CRISPRi tandem sgRNA constructs in order to identify functional constructs for subsequent experiments. I transduced the Scramble Control, MYCi and BCL2i constructs into 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells, and sorted for the brightest mCherry+/BFP+ population (to obtain successfully transduced cells with the highest dCas9-KRAB and tandem sgRNA expression levels). I then extracted RNA samples from 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells that were either untransduced (U/T), or transduced with one of the five CRISPRi constructs (Scramble Ctrl, MYCi(1), MYCi(2), BCL2i(3) or BCL2i(4)). I performed reverse transcriptase PCRs on each sample, and carried out MYC and BCL2 qPCRs to check for successful MYC and BCL2 knockdown. For each cell line system (786-M1A and OS-LM1B), I normalised the MYC and BCL2 mRNA levels in each of the transduced CRISPRi samples to the corresponding mRNA levels in the untransduced sample to determine any relative decreases in MYC or BCL2 expression.

No significant MYC knockdown was seen in cells expressing the Scramble Control or either of the BCL2-targeting constructs, BCL2i(3) and BCL2i(4), as expected (***Fig 34A***). By

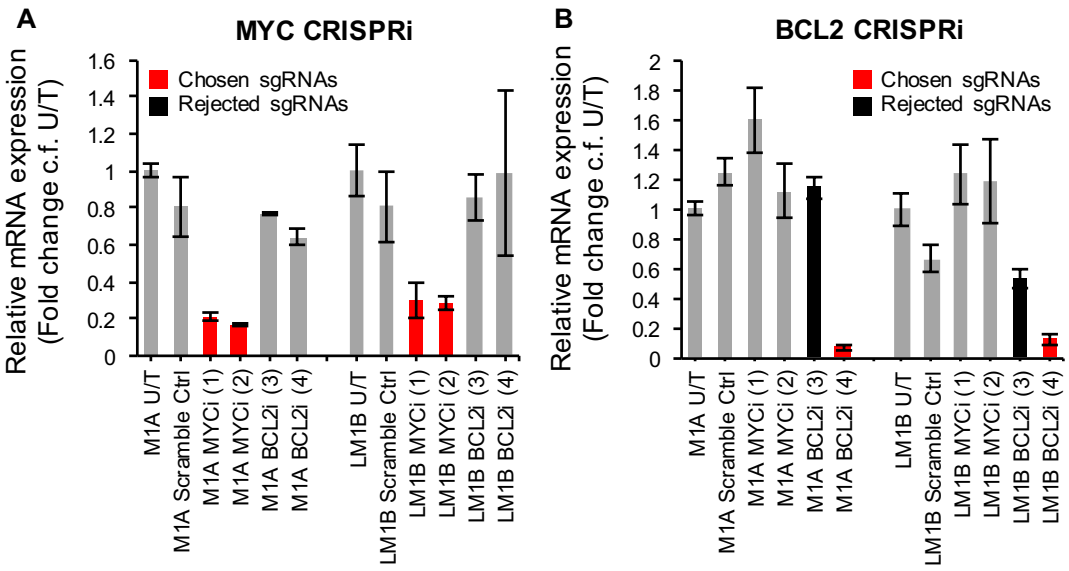


Fig 34: ***Testing of MYC and BCL2 CRISPRi tandem sgRNA constructs for optimal knockdown efficiency.***
qPCRs performed with RNA samples extracted from 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells that were either untransduced (U/T), or transduced with one of the following CRISPRi tandem sgRNA constructs: Scramble Ctrl, MYCi(1), MYCi(2), BCL2i(3) or BCL2i(4). The Scramble Ctrl comprises two non-targeting control sgRNAs. MYCi(1) and MYCi(2) are two independent MYC-targeting constructs, each comprising a different pair of tandem MYC CRISPRi sgRNAs. BCL2i(3) or BCL2i(4) are two independent BCL2-targeting constructs, each comprising a different pair of tandem BCL2 CRISPRi sgRNAs.

Tandem sgRNA constructs that successfully mediated target gene knockdown and were chosen for further experiments are highlighted in red. Tandem sgRNA constructs that failed to mediate target gene knockdown and were not chosen for further experiments are highlighted in black.

(A) MYC qPCR to check for MYC knockdown in the indicated cell lines. For each set of cell lines, the MYC mRNA levels in each sample were normalised to the Scramble Control sample. MYC knockdown was not seen in cells expressing the Scramble Control, BCL2i(3) or BCL2i(4). Both MYCi(1) and MYCi(2) mediated highly efficient knockdown of MYC.

(B) BCL2 qPCR to check for BCL2 knockdown in the same set of cell lines as *(A)*. For each set of cell lines, the BCL2 mRNA levels in each sample were normalised to the Scramble Control sample. BCL2 knockdown was not seen in cells expressing the Scramble Control, MYCi(1) or MYCi(2). The BCL2i(3) construct failed to knock down BCL2, while the BCL2i(4) construct mediated highly efficient knockdown of BCL2.

contrast, both MYC-targeting constructs, MYCi(1) and MYCi(2), mediated highly efficient knockdown of MYC. Specifically, MYCi(1) and MYCi(2) reduced MYC mRNA levels by 79% and 83.2% respectively in 786-M1A-dCas9-KRAB-mCherry CRISPRi cells, and by 70.2% and 72% respectively in OS-LM1B-dCas9-KRAB-mCherry CRISPRi cells. Thus, both these MYC CRISPRi constructs (highlighted in red in ***Fig 34A***) were selected for further experiments.

Meanwhile, BCL2 knockdown was not observed in cells expressing the Scramble Control or either of the MYC-targeting constructs, MYCi(1) and MYCi(2), again as expected (***Fig 34B***). The BCL2i(3) construct (highlighted in black in ***Fig 34B***) failed to knock down BCL2, and was thus not chosen for further experiments. However, the BCL2i(4) construct mediated highly efficient knockdown of BCL2, reducing BCL2 mRNA levels by 93.3% in 786-M1A-dCas9-KRAB-mCherry CRISPRi cells, and by 87.8% in OS-LM1B-dCas9-KRAB-mCherry CRISPRi cells. Thus, the BCL2i(4) construct (highlighted in red in ***Fig 34B***) was selected for further experiments (and shall hereafter be referred to simply as BCL2i).

To determine whether MYC and BCL2 knockdown individually (or in combination) phenocopies the selective disadvantage observed with HNF1B knockdown, I carried out CRISPRi competition proliferation assays with the functional MYCi and BCL2i constructs in both 786-M1A-dCas9-KRAB-mCherry (***Fig 35A***) and OS-LM1B-dCas9-KRAB-mCherry (***Fig 35B***) cells (hereafter referred to as 786-M1A CRISPRi and OS-LM1B CRISPRi cells respectively), using a similar protocol to that used in the HNF1Bi CRISPRi competition assay (***Fig 24A-B***).

I thus mixed untransduced parental mCherry+/BFP– cells (not transduced with sgRNAs) in a 50:50 ratio with transduced mCherry+/BFP+ cells, expressing either: Scramble Control, HNF1Bi (the HNF1B CRISPRi #2 tandem sgRNA construct that successfully knocks down HNF1B), MYCi(1) alone, MYCi(2) alone, BCL2i alone, both MYCi(1) and BCL2i, or both MYCi(2) and BCL2i. I then measured the percentages of BFP positive versus BFP negative cells every 4 days by FACs, and calculated the BFP+/BFP– ratio over time.

Consistent with the previous HNF1B CRISPRi competition assay data (***Fig 24A-B***), the BFP+/BFP– ratio in the "HNF1Bi vs untransduced" groups decreased rapidly over time in both cell lines (***Fig 35A-B***), again demonstrating the strong phenotype associated with HNF1B depletion. Meanwhile, as expected and seen before, the BFP+/BFP– ratio in the "Scramble

Control vs untransduced" groups in both cell lines remained relatively stable at approximately 1:1 (***Fig 35A-B***).

Interestingly, MYC knockdown alone, with either MYCi(1) or MYCi(2), closely phenocopied the effects of HNF1B loss. The BFP+/BFP– ratios in the "MYCi(1) vs untransduced" and "MYCi(2) vs untransduced" experimental groups also decreased rapidly over time (***Fig 35A-B***). This indicates that the untransduced MYC wild-type cells outcompeted the MYC knockdown cells in the assay mixture, suggesting that MYC loss conferred a selective disadvantage. In fact, the strength of the phenotype for MYC depletion even exceeded that seen with HNF1B knockdown. This was the case for both MYC-targeting constructs and in both cell lines, particularly the 786-M1A CRISPRi cells (***Fig 35A-B***).

Unexpectedly, however, the BFP+/BFP– ratio in the "BCL2i vs untransduced" groups phenocopied the trend in the "Scramble Control vs untransduced" groups, remaining at a ratio of approximately 1:1 over the course of each assay in both cell lines (***Fig 35A-B***). Therefore, BCL2-depleted cells grew at the same rate as the untransduced BCL2 wild-type cells in the assay mixture, and were not outcompeted. This suggests that BCL2 knockdown did not confer any selective disadvantage.

Meanwhile, the BFP+/BFP– ratios in the combination knockdown groups ("MYCi(1) + BCL2i vs untransduced", and "MYCi(2) + BCL2i vs untransduced") also declined sharply over time in both cell lines, again phenocopying the effects of HNF1B knockdown to either a similar or greater extent (***Fig 35A-B***). This indicates that cells expressing these CRISPRi constructs were also outcompeted by the untransduced cells in the assay mixture, and were thus presumably at a selective disadvantage. The magnitude of the phenotypes seen with the combination knockdowns closely resembled those seen with the individual MYC knockdowns, and did not exceed them (***Fig 35A-B***). Hence, given the lack of a phenotype with BCL2 knockdown alone, the phenotypes seen with the combination knockdowns are likely to have been entirely due to MYC loss, with little or no contribution from BCL2 depletion.

To confirm that each tandem sgRNA construct was still mediating efficient target gene knockdown in each CRISPRi competition assay, I extracted whole cell protein lysates and total RNA from each cell line, and performed qPCRs for MYC, BCL2 and HNF1B (***Fig 35C-H***).

For both the 786-M1A CRISPRi and OS-LM1B CRISPRi cell lines, I normalised the MYC, BCL2 and HNF1B mRNA levels in each transduced CRISPRi sample to the

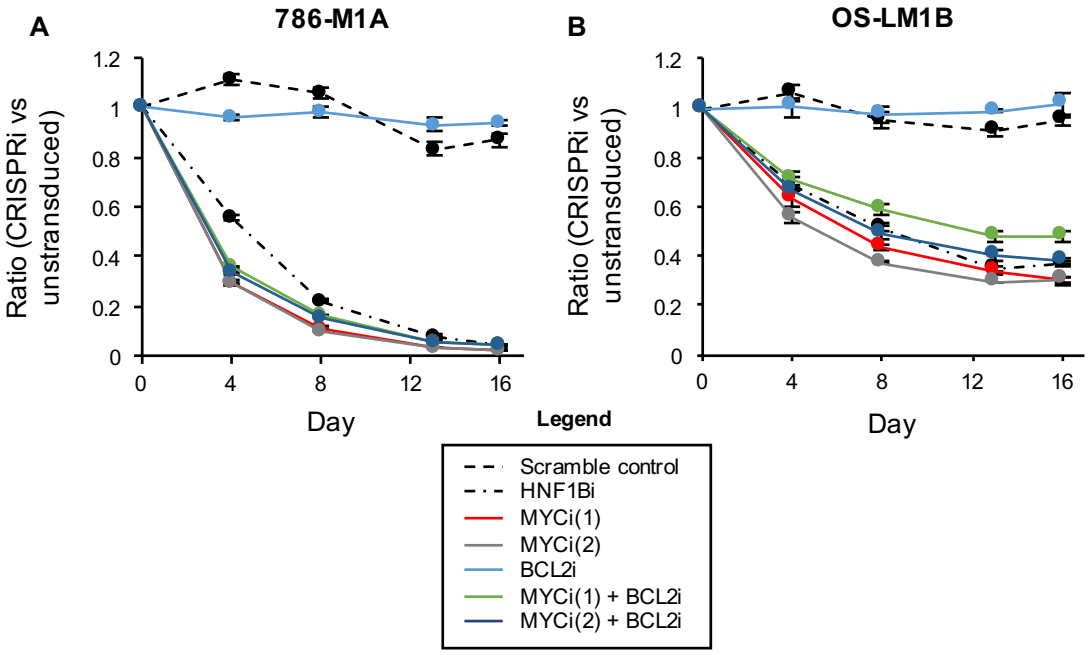


Fig 35: ***CRISPRi competition assays investigating the contributions of MYC and BCL2 to the selective disadvantage observed with HNF1B knockdown.***
(A)-(B) CRISPRi competition proliferation assays performed with the functional MYCi and BCL2i constructs, showing the effects of MYC and BCL2 knockdown (individually and in combination) relative to the effects of HNF1B knockdown in (*A*) 786-M1A-dCas9-KRAB-mCherry and (*B*) OS-LM1B-dCas9-KRAB-mCherry cells. In each assay, untransduced BFP– cells (not containing sgRNAs) were mixed in 50:50 ratios with transduced BFP+ cells expressing one of the indicated CRISPRi tandem sgRNA constructs (*see Legend*). The percentages of BFP positive versus BFP negative cells in each cell mixture were measured every 4 days by FACs, and used to calculate the BFP+/BFP– ratio over time. Each graph shows the changes in this BFP+/BFP– ratio over the course of each assay, for each cell mixture (i.e. each experimental group).

(C)-(H) qPCRs for (*C-D*) MYC, (*E-F*) BCL2 and (*G-H*) HNF1B in the (*C, E, G*) 786-M1A-dCas9-KRAB-mCherry cell lines and (*D, F, H*) OS-LM1B-dCas9-KRAB-mCherry cell lines used in the MYCi / BCL2i competition assays. The MYC, BCL2 and HNF1B mRNA levels in each transduced CRISPRi sample were normalised to the corresponding mRNA levels in the untransduced cell samples for each cell line system to check for successful target gene knockdown. The HNF1Bi, MYCi and BCL2i constructs all successfully mediated HNF1B, MYC and BCL2 knockdown respectively in both competition assays.

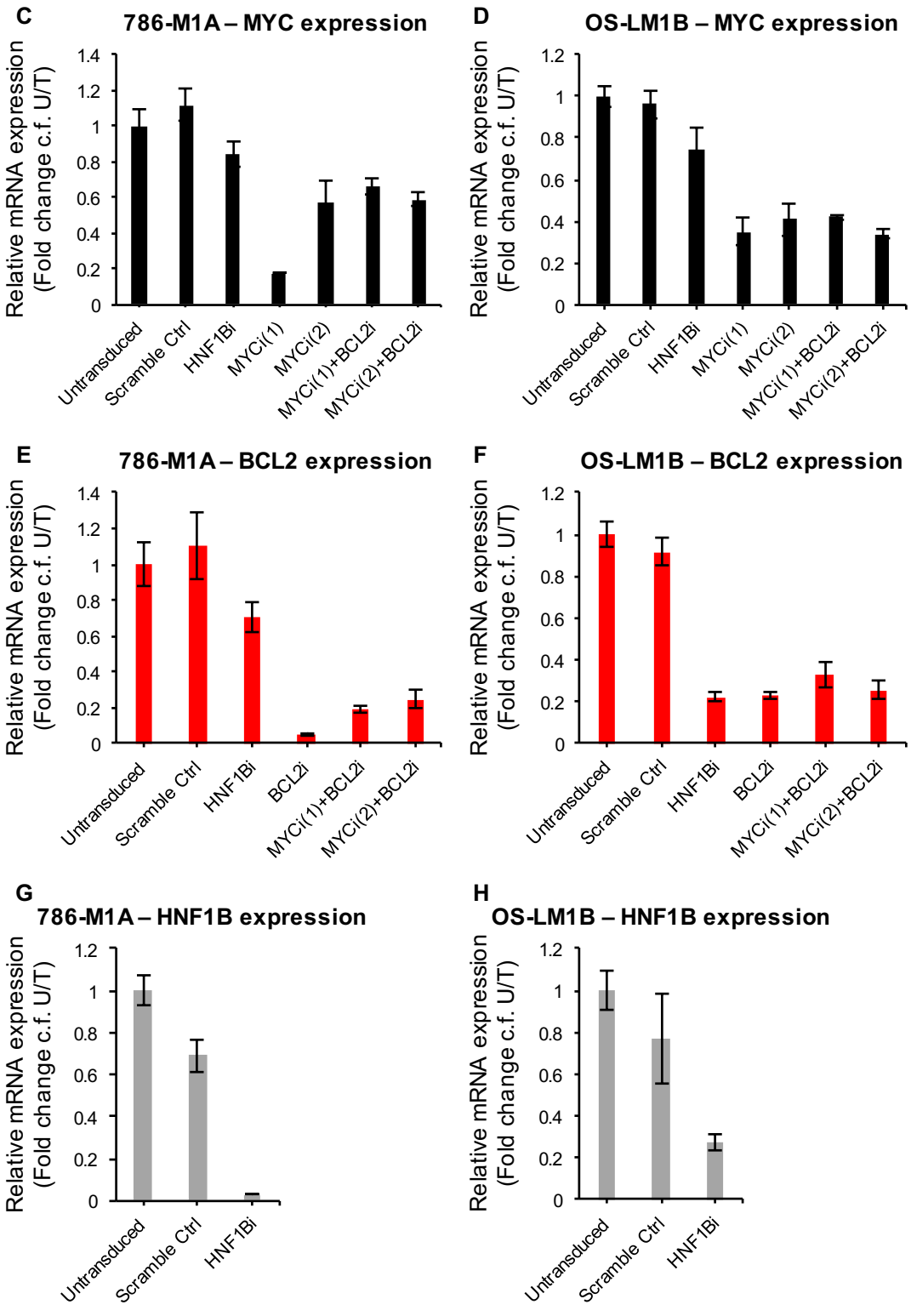


Fig 35 (continued): ***CRISPRi competition assays investigating the contributions of MYC and BCL2 to the selective disadvantage observed with HNF1B knockdown.***

corresponding mRNA levels in the untransduced sample to check for successful target gene knockdown.

In the MYC qPCR data (***Fig 35C-D***), MYC levels did not decrease in 786-M1A CRISPRi or OS-LM1B CRISPRi cells transduced with the Scramble Control, as expected. MYC expression was reduced in cells expressing the HNF1Bi construct, with a fold change of 0.840 (i.e. a 16% decrease) in the 786-M1A CRISPRi cells (***Fig 35C***) and a fold change of 0.742 (i.e. a 25.8% decrease) in the OS-LM1B CRISPRi cells (***Fig 35D***).

MYC was successfully knocked down by the MYCi(1) and MYCi(2) constructs, both in cells expressing those constructs alone, and in cells that also expressed BCL2i. In the 786-M1A CRISPRi cells, the MYCi(1) construct alone gave the most efficient MYC knockdown, reducing MYC levels by 82.9% (a fold change of 0.171) (***Fig 35C***). Meanwhile MYCi(2) alone, MYCi(1) in combination with BCL2i, and MYCi(2) in combination with BCL2i all mediated efficient MYC knockdown to a roughly equal extent, inducing 42.8%, 34.3% and 41.2% decreases in MYC expression respectively (i.e. fold changes of 0.572, 0.657 and 0.588 respectively) (***Fig 35C***). In the OS-LM1B CRISPRi cells, MYCi(1) and MYCi(2), either alone or with the additional presence of BCL2i, gave similarly efficient MYC knockdowns (***Fig 35D***). Cells expressing only MYCi(1) or MYCi(2) showed MYC fold changes of 0.351 or 0.410 (i.e. 64.9% or 59% decreases) respectively (***Fig 35D***). Meanwhile, cells expressing "MYCi(1) + BCL2i" or "MYCi(2) + BCL2i" showed MYC fold changes of 0.419 or 0.341 (i.e. 58.1% or 65.9% decreases) respectively (***Fig 35D***). Therefore, these results confirmed that the MYC tandem sgRNA constructs were functional in the cell lines used in both CRISPRi competition assays.

Meanwhile, in the BCL2 qPCR data (***Fig 35E-F***), BCL2 levels did not decrease in cells expressing the Scramble Control, but they did fall in cells expressing HNF1Bi, with a fold decrease of 0.705 (a 29.5% reduction) in 786-M1A CRISPRi cells (***Fig 35E***), and a fold decrease of 0.220 (a 78% reduction) in OS-LM1B CRISPRi cells (***Fig 35F***).

The BCL2i construct mediated highly efficient BCL2 knockdown in each cell line, both when expressed alone and when expressed in cells that also contained either of the MYC-targeting constructs (***Fig 35E-F***). In the 786-M1A CRISPRi cells, BCL2 mRNA levels decreased by 94.9% (a fold change of 0.051) in cells expressing the BCL2i construct alone, by 80.9% (a fold change of 0.191) in cells expressing "MYCi(1) + BCL2i", and by 75.4% (a fold change of 0.246) in cells expressing "MYCi(2) + BCL2i" (***Fig 35E***). In the OS-LM1B CRISPRi

cells, the extent of BCL2 knockdown was also fairly similar between the three BCL2i construct combinations, and also similar to the fall in BCL2 levels in HNF1Bi-expressing cells (***Fig 35F***). BCL2 mRNA levels were reduced by 76.9% (a fold change of 0.231) in cells expressing BCL2i alone, by 67.5% (a fold change of 0.325) in cells expressing "MYCi(1) + BCL2i", and by 74.4% (a fold change of 0.256) in cells expressing "MYCi(2) + BCL2i" (***Fig 35F***).

These results verify that the BCL2i tandem sgRNA construct was indeed functional in the cell lines used in both CRISPRi competition assays. Therefore, the absence of an observable phenotype with individual BCL2 knockdown was not because the BCL2i tandem sgRNA construct failed to knock down BCL2.

Finally, in the HNF1B qPCR data (***Fig 35G-H***), HNF1B levels were considerably depleted in cells expressing the HNF1Bi construct (a.k.a HNF1B CRISPRi #2) (***Fig 35G-H***), as expected based on its previous use in the HNF1B CRISPRi competition assay (***Fig 24***). In the 786-M1A CRISPRi cells transduced with HNF1Bi, there was a fold decrease in HNF1B expression of 0.029 (i.e. a 97.1% reduction) (***Fig 35G***). In the OS-LM1B CRISPRi cells transduced with HNF1Bi, there was a fold decrease in HNF1B expression of 0.276 (i.e. a 72.4% reduction) (***Fig 35H***). These values are consistent with the extent of HNF1B knockdown seen in the cell lines used for the HNF1B CRISPRi competition assay (***Fig 24C***).

In summary, MYC knockdown phenocopies the selective disadvantage observed with HNF1B depletion, producing a similarly strong (or even stronger) phenotype, whereas BCL2 knockdown has no apparent effect. This was observed in competition assays in both 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells, thus indicating that the equally striking phenotype conferred by MYC depletion is a reproducible result, as it was seen in multiple ccRCC cell line systems. The HNF1Bi, MYCi and BCL2i constructs all successfully mediated HNF1B, MYC and BCL2 knockdown respectively in these MYCi / BCL2i competition assays. Therefore, the phenotypes observed in these experiments are likely to be the results of target gene knockdown.

3.7) MYC, and possibly BCL2, may be involved in the phenotype seen with HNF1B loss

The absence of a phenotype with BCL2 knockdown alone was a somewhat unexpected result. The data from the Annexin V/PI stainings had clearly demonstrated a marked increase in apoptosis (specifically late-stage apoptosis and/or necrosis) in HNF1B knockout cells, which was

rescued by HNF1B restoration, thus suggesting that the increased cell death was linked to HNF1B loss. This apparent increase in apoptosis highlighted the possibility that loss of the anti-apoptotic protein BCL2 contributed to the selective disadvantage associated with HNF1B depletion.

One possibility is that apoptosis may not be a contributing factor to the selective disadvantage following HNF1B loss, and instead impaired cell cycle progression (and thus inhibited ccRCC cell proliferation) plays more of a role. However, it is also possible that HNF1B depletion might induce an alternative cell death pathway, one that might not involve BCL2 (such as necrosis). Alternatively, HNF1B depletion might induce increased apoptosis via a different mechanism (and not via the downregulation of BCL2) or with the involvement of additional factors rather than just BCL2 alone. If any of these possibilities are true, then this might explain the inability of BCL2 knockdown alone to even partially phenocopy the effects of HNF1B loss.

Another explanation might be that the increased cell death observed in HNF1B knockout cells is related to the downregulation of MYC, with the downregulation of BCL2 serving to render the knockout cells more susceptible to apoptosis. This explanation would be consistent with the data from the Annexin V/PI apoptosis assays suggesting that enhanced cell death contributes to the selective disadvantage associated with HNF1B depletion, and also consistent with the lack of a phenotype seen with individual BCL2 knockdown. Additionally, this explanation is supported by the data showing the existence of a cell cycle defect in HNF1B knockout cells, and the data demonstrating the ability of MYC knockdown alone to phenocopy the selective disadvantage seen with HNF1B loss.

These ideas are developed further in the Discussion section of this thesis. Altogether, my data suggests the possibility that MYC might be primarily involved in the mechanisms by which HNF1B supports ccRCC, with secondary involvement of BCL2. Ultimately, HNF1B's ability to regulate MYC might at least partially explain the importance of HNF1B in ccRCC.

DISCUSSION

1) Screening for transcription factor dependencies in ccRCC

ccRCC is often highly aggressive and prone to metastasis, and exhibits substantial genetic intratumour heterogeneity that may contribute to the frequent development of drug resistance in these tumours in response to the current standard-of-care targeted systemic therapies (such as kinase inhibitors, anti-angiogenic drugs and mTOR inhibitors[1,2,32–41,24,42–46,25–31]. To effectively combat ccRCC, novel therapeutic targets are required, ideally those that would be essential to ccRCC cell survival and growth across all tumour subclones.

We reasoned that transcription factors are likely to be ubiquitously essential, given their important roles in determining and regulating cell phenotypes. Moreover, transcription factors function as effectors of multiple signalling pathways. Hence, therapeutic perturbation of a single transcription factor could have wide-ranging effects on ccRCC cell fitness and survival, compared to targeting signalling components further upstream in a given pathway. Additionally, there are limited opportunities for tumours to rely on redundancy to circumvent transcription factor targeting. There is substantial evidence in the literature demonstrating that various transcription factors play important roles in many different tumour types (often in cases where lineage-specific transcription factors seem to be essential in tumours derived from the tissue type in question), and that some of these key transcription factors are in fact druggable and thus have considerable potential as therapeutic targets.

The focus of my PhD was thus to identify additional and potentially novel ccRCC cell dependencies, and characterise their importance in these tumours. To that end, I conducted a CRISPR/Cas9-based genetic depletion screen *in vitro*, using ccRCC cells expressing a dox-inducible Cas9 transgene (786-M1A-TGL-tet-on-Cas9 cells), with a lentiviral sgRNA library targeting 50 RCC-associated transcription factors that were particularly highly expressed in ccRCC (***Fig 1*** to ***Fig 7***).

This screen revealed a potential role for the renal developmental transcription factor HNF1B in supporting ccRCC. Apart from HIF2A (which is already known to be important in ccRCC), HNF1B was the strongest hit, dropping out the most significantly out of the remaining 49 transcription factor genes: all 8 individual HNF1B sgRNAs from the screen were

substantially depleted by the end of the experiment (all falling within the top 50 most depleted sgRNAs out of the 420 constructs in the library), and HNF1B showed the strongest median depletion score with the second smallest p-value (***Fig 8***, ***Fig 9***).

Interestingly, the 20 non-targeting control sgRNAs in the screen often appeared relatively enriched when comparing the sgRNA representations between screen samples (***Fig 9B-D***), but to varying extents. The "enrichment score" observed varied between the 20 non-targeting control sgRNAs in each sample comparison, and the number of non-targeting controls that appear relatively enriched also varied between sample comparisons (***Fig 9B-D***).

In the comparison of Day 26 +dox to Baseline Day 0, all 20 non-targeting controls are enriched (***Fig 9B***). In the comparison of Day 26 +dox to Day 26 –dox, 19 non-targeting controls appear enriched, with one appearing to be neither depleted nor enriched (showing only a very high depletion score of 0.974), (***Fig 9C***). Meanwhile, in the comparison of Day 26 –dox to Baseline Day 0, 5 control sgRNAs appear to be similarly barely depleted, with only 15 out of 20 non-targeting constructs appearing relatively enriched (***Fig 9D***).

This could be due to a number of possible reasons. Perhaps the most likely reason is that cells containing non-targeting control sgRNAs would not have experienced double-stranded break (DSB) formation, due to the lack of Cas9 targeting. Without this source of cellular stress from DNA cleavage and repair, these cells may have been at a slight selective advantage compared to cells expressing targeting sgRNAs, even if those sgRNAs targeted genes that did not influence ccRCC cell fitness. Thus, cells expressing non-targeting control sgRNAs may have proliferated better and outcompeted other transduced cells, resulting in the non-targeting control sgRNAs appearing relatively enriched.

Indeed, consistent with this, the non-targeting controls appear more enriched (both in terms of their "enrichment score" and the number of enriched versus unchanged / mildly depleted control sgRNAs) when comparing Day 26 +dox to Baseline Day 0 (***Fig 9B***) and Day 26 +dox to Day 26 –dox (***Fig 9C***), than they do when comparing Day 26 –dox to Baseline Day 0 (***Fig 9D***).

In the dox-treated sample, cells expressing control sgRNAs would be directly competing against cells expressing targeting sgRNAs (i.e. cells experiencing DSB formation). Thus, the selective advantage for cells not experiencing Cas9 activity and cleavage would be more pronounced, allowing them to better outcompete the other sgRNA-transduced cells in the dox-

treated population, and resulting in the non-targeting control sgRNAs becoming relatively overrepresented (i.e. relatively enriched) in the final sgRNA pool for the Day 26 +dox sample. In the untreated sample, however, cells containing targeting sgRNAs should also not be experiencing Cas9 activity and DSB formation in the absence of dox. Thus, barring any leakiness of Cas9 activity, cells expressing control sgRNAs and cells expressing targeting sgRNAs should proliferate equally, as the former would have relatively less selective advantage. Hence, the representation of the non-targeting controls would be more equal to that of the targeting sgRNAs in the final genomic DNA pool of the Day 26 –dox sample, and the non-targeting controls would appear less relatively enriched. In this way, the varying enrichment of non-targeting controls could have occurred in the absence of any off-target effects by the non-targeting sgRNAs themselves.

Of course, it is entirely possible that at least some of the control sgRNAs may have had off-target effects that led to increased cell proliferation in cells that expressed them. Alternatively, control sgRNA constructs may have become integrated into the genome during lentiviral transduction at sites which enhanced the effects of genes that promote proliferation, or dampened the effects of genes that limit proliferation.

It is also possible that the apparent enrichment of the non-targeting control sgRNAs was simply due to the fact that there were a lot of real "hits" in the target gene set whose sgRNAs appeared depleted at the end of the screen. As a result, the sgRNAs targeting genes that were not "hits", including the non-targeting control sgRNAs, would appear enriched by comparison. Indeed, the 50 transcription factor target genes in the screen library were a very highly selected group of genes, all chosen because they were more likely to be important in ccRCC based on their high expression in this tumour type.

I next confirmed the knockout efficiency of each of the 8 HNF1B sgRNAs (relative to the NTC18 non-targeting control sgRNA) by transducing each construct individually into fresh 786-M1A-TGL-tet-on-Cas9 cells, treating the cells with dox for 8 days, and blotting for HNF1B in whole cell protein lysates extracted from these nine cell lines. Most of the constructs were highly effective at knocking out or knocking down HNF1B (***Fig 10A***); however, there were discrepancies between the extent of HNF1B protein loss achieved versus the depletion score observed in the screen for each HNF1B sgRNA (***Fig 9E***, ***Fig 10A-B***).

These discrepancies between depletion score and knockout efficiency may be due to off-target effects of the sgRNAs, in cases where the construct dropped out to a greater extent than its ability to mediate HNF1B protein loss.

Alternatively, experimental differences may be responsible: sgRNA knockout efficiency was tested after only 8 days of dox treatment, whereas sgRNA depletion was assessed after 26 days on dox. Thus, an sgRNA that dropped out considerably in the screen might in fact mediate more complete protein knockout over the course of 26 days compared to what I saw after only a week of dox treatment. On the other hand, 26 days of dox treatment may provide more opportunity to select for resistant or non-knockout cells expressing a particularly effective sgRNA. Such escapers might survive and proliferate, misleadingly increasing the representation of that sgRNA in the final genomic DNA pool. That sgRNA may thus seem less depleted in the screen, while still exhibiting a high knockout efficiency in the Western blot data after a week of dox treatment.

Of course, technical error (such as HNF1B protein degradation while extracting the protein lysates, or inefficient PCR amplification of certain sgRNA constructs from the screen genomic DNA samples) could have affected the data for sgRNA knockout efficiency or for relative sgRNA abundance (and thus sgRNA depletion score), thus explaining the inconsistent results. Another explanation for the discrepancies between depletion score and knockout efficiency is that they could simply be due to random noise in the screen data that affected the sgRNA depletion scores obtained.

I selected HNF1B sgRNAs #2 and #8 (sg2 and sg8 respectively) for further experiments, as both mediated highly efficient HNF1B protein knockout, but sg2 dropped out to a lesser extent in the screen relative to sg8. I was thus able to compare the results observed between the two sgRNAs in my validation experiments to determine if any off-target effects (indicated by different or stronger phenotypes) were associated with the more depleted sgRNA, sg8. However, both constructs ultimately exhibited highly consistent phenotypes, thus strengthening the conclusions from my data (as they were validated with two independent HNF1B sgRNAs).

2) Establishing a phenotype for HNF1B CRISPR/Cas9 knockout in vitro

I demonstrated a reproducible phenotype for HNF1B knockout *in vitro* with both sg2 and sg8, via CellTitre Glo, cell counting and competition proliferation assays, in multiple ccRCC cell lines (each expressing a constitutive Cas9 transgene; ***Fig 13*** to ***Fig 19***). The data from these assays clearly established that there was a striking selective disadvantage associated with HNF1B knockout with either sg2 or sg8, relative to control cells expressing NTC18. Western blotting for HNF1B in the ccRCC-Cas9 cell lines used for these proliferation assays confirmed that HNF1B protein was successfully knocked out by each sgRNA in each experiment (***Fig 20***).

I did not observe this selective disadvantage in HeLa-Cas9 cells that were transduced with sg2 or sg8. Since HeLas do not express HNF1B, the absence of a phenotype suggested that the phenotype seen with HNF1B knockout in ccRCC cells was not the result of off-target effects caused by the sgRNAs (***Fig 21***, ***Fig 22***).

Moreover, I showed that the selective disadvantage observed in ccRCC cells was rescued by the expression of an exogenous HNF1B transgene (modified to be untargetable by sg8), suggesting that the phenotype observed was indeed a consequence of HNF1B loss (***Fig 23A-D***). Western blotting for HNF1B in the cell lines used for these rescue competition assays confirmed that the exogenous HNF1B construct was successfully expressed (***Fig 23E-F***). In retrospect, it would also have been good to blot for the 6xMyc tag as well (attached to the N-terminus of the exogenous HNF1B protein), to further confirm transgene expression.

In the 786-M1A-Cas9 BFP and mCherry cell lines, the levels of exogenous HNF1B expressed in the sg8 (+ rescue) cells were similar to the levels of endogenous HNF1B expressed in the Control (+ empty vector) cells (***Fig 23E-F, lanes 1-3 in each***). In the OS-LM1B-Cas9 BFP and mCherry cell lines, the levels of exogenous HNF1B expressed in the sg8 (+ rescue) cells were higher than the levels of endogenous HNF1B expressed in the Control (+ empty vector) cells (***Fig 23E-F, lanes 4-6 in each***), but potentially without reaching levels of overexpression that would introduce artifacts in the data. This is suggested by the fact that the extent of rescue was similar at the Day 12 time point between both sets of 786-M1A-Cas9 and OS-LM1B-Cas9 rescue competition assays (***Fig 23A-D***).

It is difficult to predict whether the extent of rescue in the OS-LM1B-Cas9 rescue competition assays would have become less complete over time, to match the extent of rescue seen in the 786-M1A-Cas9 rescue competition assays. The OS-LM1B-Cas9 rescue competition

assays were carried out for a shorter time period (12 days, versus 28 days for the 786-M1A-Cas9 assays) due to the increased sickliness of the OS-LM1B-Cas9 sg8 (+ empty vector) cell lines. As mentioned previously in the Results section, the selective disadvantage associated with HNF1B loss was more pronounced in OS-LM1B-Cas9 knockout cells relative to 786-M1A-Cas9 knockout cells (evident from the competition assay data in ***Figures 14A-B***, ***15A-B***, ***16A-B***, ***17A-B***, and ***23A-D***). If I had had more time on this project, I would have repeated these rescue competition assays to confirm how well the rescue phenotype was maintained over time in both the 786-M1A-Cas9 and OS-LM1B-Cas9 sets of cell lines.

One possible explanation for the somewhat incomplete rescue seen in the 786-M1A-Cas9 rescue competition assays is that not all cells within the population might express exogenous HNF1B at the levels observed by Western blot. The protein expression detected by Western blot only represents the population-level expression of HNF1B; however, in reality there would be more of a distribution in the levels of exogenous HNF1B expression, from cells that express very low levels to cells that express very high levels of the HNF1B rescue construct. Thus, if too few cells in the population express close to endogenous levels of exogenous HNF1B, the extent of rescue seen might be lower than expected.

The slight overexpression of exogenous HNF1B in the OS-LM1B-Cas9 sg8 (+ rescue) cell lines was unexpected, as I did initially perform virus titration experiments (*data not shown*) to optimise the amount of HNF1B rescue construct virus to use to transduce parental 786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cells for these rescue competition assays. Specifically, I titrated the virus by infecting parental 786-M1A-Cas9 cells with different amounts of virus for the HNF1B rescue construct, extracted whole cell protein lysates from these transduced cells, and blotted for HNF1B to see which virus quantity produced exogenous HNF1B expression levels similar to the expression levels of endogenous HNF1B. I then used this volume of virus to transduce fresh 786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cells with either empty vector or the exogenous HNF1B rescue construct, and then transduced these cell lines with the Control and sg8 constructs and carrying out the rescue competition assays.

Based on my virus titration experiments, I expected the sg8 (+ rescue) cell lines used in the competition assays to exhibit exogenous HNF1B levels that were similar to the levels of endogenous HNF1B, and this was indeed the case for the 786-M1A-Cas9 sg8 (+ rescue) cells, but not for the OS-LM1B-Cas9 sg8 (+ rescue) cells. The reasons for this difference are unknown,

but could be related to the different transduction efficiencies of different cell lines: the OS-LM1B-Cas9 cells may simply have been better at taking up virus for the HNF1B rescue construct, and this expressed higher levels of exogenous HNF1B.

If I had had more time on this project, I would have tried to employ strategies to modulate the expression levels of the exogenous HNF1B transgene. For instance, I could have tried to modify the promoter of the construct, perhaps by testing different promoter truncations, to try to reduce transgene expression to match endogenous levels. I could then have transduced this modified construct into 786-M1A-Cas9 and OS-LM1B-Cas9 parental cells, and then transduced these new cell lines with sgRNAs and repeated the rescue competition assays.

Alternatively, I could have transduced the 786-M1A-Cas9 and OS-LM1B-Cas9 parental cells with the exogenous HNF1B rescue construct, and then sorted for single cell clones. I could then have screened for HNF1B-restored single cell clones that expressed levels of exogenous HNF1B similar to the levels of endogenous HNF1B expressed in wild-type control cells. I could either select a number of individual clones to use as biological replicates, or pool several clones together, to try to account for potential issues arising from clonal differences.

3) Establishing a phenotype for HNF1B CRISPRi knockdown in vitro

I next tested the effects of CRISPRi-mediated HNF1B knockdown in ccRCC cell lines, and observed a similarly strong selective disadvantage to that seen with CRISPR/Cas9-mediated HNF1B knockout (***Fig 24A-B***), thus suggesting that the phenotype seen with CRISPR/Cas9 HNF1B knockout was not due to a loss of cell fitness associated with double-stranded break formation by Cas9.

However, the extent of HNF1B knockdown was considerably less in OS-LM1B-dCas9-KRAB-mCherry cells expressing the HNF1B CRISPRi #2 construct, compared to that achieved in 786-M1A-dCas9-KRAB-mCherry cells expressing HNF1B CRISPRi #2. This was evident by both HNF1B qPCR (***Fig 24C***) and HNF1B Western blot (***Fig 24D***).

This could be due to the fact that expression of the dCas9-KRAB-mCherry transgene was less stable over time in OS-LM1B cells relative to 786-M1A cells (*data not shown*). The OS-LM1B-dCas9-KRAB-mCherry cells gradually lost mCherry expression over time, even in the absence of tandem sgRNA constructs, and separated out into an mCherry+ or mCherry^{high}

population (expressing high levels of mCherry) and a considerably downwards shifted mCherry^{low}, or more often an mCherry–, population (expressing very low levels of mCherry at best, or none at all) (*data not shown*). Since mCherry is a marker of dCas9-KRAB expression, this gradual and spontaneous loss of mCherry fluorescence suggested a loss of dCas9-KRAB expression over time in the OS-LM1B transduced population. Consequently, the efficiency of target gene knockdown would be greatly reduced, leading to higher residual levels of target gene expression. (***Fig 24C-D***).

While this phenomenon was also observed in 786-M1A-dCas9-KRAB-mCherry cells, it was to a much lesser extent. At worst, the whole population only separated into an mCherry^{high} and a slightly downwards shifted mCherry^{low} population (*data not shown*). Thus, the CRISPRi knockdown efficiency achieved in these cells was much higher (***Fig 24C-D***).

Notably, even the partial decrease in HNF1B expression seen in the OS-LM1B-dCas9-KRAB-mCherry cells (***Fig 24C-D***) was sufficient to induce a striking selective disadvantage (***Fig 24B***), similar in strength to the selective disadvantage seen with more complete HNF1B knockdown in 786-M1A-dCas9-KRAB-mCherry cells (***Fig 24A***), and with efficient HNF1B knockout in the 786-M1A-Cas9 and OS-LM1B-Cas9 cells with CRISPR/Cas9 (***Figs 14A-B***, ***15A-B***, ***16A-B***, ***17A-B***, and ***23A-D***). This further underscores the importance of HNF1B in ccRCC.

4) Establishing a phenotype for HNF1B CRISPR/Cas9 knockout in vivo

I then conducted subcutaneous tumour xenograft assays to determine whether HNF1B loss would have an impact on ccRCC cells *in vivo*, and indeed HNF1B knockout also seemed to impair tumour growth. Tumour growth was substantially reduced in 786-M1A-Cas9 and OS-LM1B-Cas9 sg2 and sg8 knockout tumours relative to Control tumours, based on bioluminescence imaging (***Fig 25A-B***), calliper measurements of tumour volume (***Fig 25C-D***), and comparisons of the final tumour weights between control and knockout tumours at the end of each experiment (***Fig 25E-F***).

Prior to subcutaneous injection, I confirmed that the Control cell lines used were HNF1B wild-type, and that the sg2 and sg8 cell lines were HNF1B knockout, by HNF1B Western blot (***Fig 26A***). At the end of each experiment, I performed HNF1B immunohistochemistry (***Fig 26B,***

D) to determine whether the knockout tumours were comprised of HNF1B-positive or HNF1B-negative cells (with the control tumours as a comparison point), in order to deduce whether the small sg2 and sg8 tumours that formed were comprised of cells that had escaped knockout and still expressed HNF1B, or cells that were HNF1B knockout but subsequently adapted to HNF1B loss. I also performed vimentin immunohistochemistry (***Fig 26C, E***) with sections from the same tumour samples to confirm that the control and knockout tumours were comprised of human ccRCC cells (vimentin-positive), and determine whether they contained murine cells (vimentin-negative) from the surrounding mouse stromal tissue. The HNF1B and vimentin IHC revealed that both the control and knockout tumours contained a mixture of HNF1B-positive and HNF1B-negative cells, and a mixture of vimentin-positive and vimentin-negative cells (***Fig 26B-E***).

The Control tumours contained a higher proportion of more intensely staining HNF1B-positive nuclei and a minority of HNF1B-negative nuclei, as well as exhibiting both vimentin-positive and vimentin-negative cells (***Fig 26B-E; left image in each panel***). Since these tumours still expressed wild-type HNF1B and exhibit rapid growth *in vivo*, these IHC results could indicate that the Control tumours incorporated murine stromal cells as they grew, for instance through the recruitment of blood vessels and lymphatic vessels. Other stromal components may also have been present, such as fibroblasts and immune cells, which would be derived from the surrounding mouse tissue. Thus, the small proportion of HNF1B-negative nuclei in the Control tumours may have corresponded to vimentin-negative non-ccRCC cells in the tumour mass.

Meanwhile, the sg2 and sg8 tumours contained a greater proportion of HNF1B-negative nuclei, and a smaller proportion of HNF1B-positive nuclei that stained less intensely, initially suggesting that these tumours simply consisted of a mixture of both non-knockout and HNF1B-independent ccRCC cells (***Fig 26B, D; middle & right images in each panel***). However, the presence of both vimentin-positive and vimentin-negative cells in the knockout tumours suggested that they had also incorporated murine, non-ccRCC cells into the tumour mass (***Fig 26C, E; middle & right images in each panel***), despite supposedly being at a selective disadvantage and exhibiting substantially reduced growth *in vivo*.

One possibility is that the knockout tumours were derived from sg2- and sg8-expressing cells that had escaped HNF1B knockout or become resistant to HNF1B loss, and that these escaper cells may have still been able to grow after overcoming the effects of the sgRNAs. These escaper cells might thus have been in the early stages of forming tumours by the endpoint of

each subcutaneous tumour xenograft assay, when the small "knockout" tumours were harvested. This could explain why each of the small sgRNA-expressing tumours comprised both vimentin-positive ccRCC cells and vimentin-negative murine cells, and suggests that the HNF1B-negative nuclei may have corresponded to both vimentin-positive, HNF1B-depleted ccRCC cells that had adapted to HNF1B loss, and vimentin-negative non-ccRCC cells derived from mouse stromal tissue that was incorporated as the tumours began to grow.

However, this is all speculation, as I was not able to identify matched regions between the HNF1B-stained and vimentin-stained tumour sections for each of the Control, sg2, and sg8 tumours analysed. I was thus unable to directly compare the same tumour regions for the HNF1B and vimentin IHC to determine whether HNF1B +/- cells exactly corresponded to vimentin +/- cells. Ideally, I would have used a programme, such as the Halo software (*Indica Labs*), to analyse the scanned images of my HNF1B and vimentin stained tumour sections. I would have then been able to compare the same fields of view for each tumour in order to match HNF1B +/- cells to the vimentin +/- cells in the same region.

In addition, different tumour samples stained for HNF1B and vimentin with different intensities, despite using the same antibody concentration; thus, if I had had more time, I would have further optimised the HNF1B and vimentin staining protocols to try to achieve more uniform staining intensities and avoid overstaining of any tumour samples or regions within samples.

Moreover, if I had had more time, I would also have carried out additional IHC stainings to further confirm the phenotypes associated with HNF1B loss *in vivo*. I would have stained for Ki67, a marker for proliferation, in order to confirm that ccRCC cells in the control tumours were indeed proliferating more than ccRCC cells in the sg2 and sg8 tumours. I would also have stained for MYC and BCL2 expression, as these two key genes were downregulated upon HNF1B knockout *in vitro* based on RNA-seq data (*see Discussion Section 6 below*). It would thus have been useful to confirm whether MYC and BCL2 expression also decreased in HNF1B knockout tumours.

5) Future work: additional experiments to investigate the effects of HNF1B loss in vivo

Due to limited time, I was only able to conduct one *in vivo* experiment (but with two cell line systems) testing the effects of constitutive HNF1B knockout on tumour growth (***Fig 25***). This *in vivo* data thus complemented the *in vitro* competition and proliferation assay data obtained with the constitutive Cas9 cell lines, in which HNF1B was also constitutively knocked out, and the proliferation phenotype of constitutive knockout cells was compared against that of control cell lines (***Figs 14 to 19***).

If I had had more time, I would also have carried out a subcutaneous tumour xenograft assay with constitutive HNF1B rescue, to complement the *in vitro* data obtained from the rescue competition assay experiments (***Fig 23***). I would have injected freshly transduced 786-M1A-Cas9 and OS-LM1B-Cas9 Control (+ empty vector), sg8 (+ empty vector), and sg8 (+ rescue) cell lines, in order to compare the growth of control tumours (comprised of cells transduced with NTC18 and empty vector, which thus express wild-type endogenous HNF1B), HNF1B knockout tumours without rescue (comprised of cells transduced with HNF1B sg8 and empty vector, which thus lack HNF1B expression entirely), and HNF1B knockout tumours *with* rescue (comprised of cells transduced with HNF1B sg8 and the constitutive HNF1B rescue construct, which thus lack endogenous HNF1B but express exogenous HNF1B).

This constitutive HNF1B rescue experiment would then have enabled me to determine whether the selective disadvantage seen with constitutive HNF1B knockout could also be rescued by HNF1B restoration *in vivo*. If so, this would indicate that the reduced growth of knockout tumours is linked to HNF1B loss in the tumour cells, and would further demonstrate the importance of HNF1B *in vivo* as well as *in vitro*.

It should also be noted that, based solely on the constitutive HNF1B knockout experiment that I carried out (***Fig 25***), the exact stage of tumourigenesis in which HNF1B is important is currently uncertain. HNF1B might be essential for tumour initiation, and thus HNF1B loss from the point of tumourigenesis would ultimately result in smaller endpoint tumours. Alternatively, HNF1B might be critical for tumour maintenance in established tumours (i.e. in the later stages of ccRCC development). The selective disadvantage that resulted from constitutive HNF1B knockout, with an absence of HNF1B from the very beginning of the experiment, does not exclude either possibility: the reduced tumour growth observed could indicate a dependence on HNF1B for tumour initiation or a dependence on HNF1B for tumour maintenance.

To try to discern which stage of tumour development HNF1B is required for, it would have been informative to also conduct *in vivo* experiments testing the effects of HNF1B depletion in established tumours.

One option would have been to carry out subcutaneous tumour xenograft assays with conditional HNF1B knockout instead of constitutive, in order to induce HNF1B loss at a later point, once tumours had already formed. This could have been done using the 786-M1A-TGL-tet-on-Cas9 clonal cell line originally used for the *in vitro* CRISPR/Cas9 genetic depletion screen (*data not shown*). These cells express Cas9 under the control of a dox-inducible promoter. Once transduced with constitutively expressed HNF1B sgRNAs and injected subcutaneously, these cells could be used to achieve conditional HNF1B knockout *in vivo* at different time points, with the administration of doxycycline via the mouse food pellets.

I could thus have used 786-M1A-TGL-tet-on-Cas9 clonal cells transduced with the Control (NTC18), HNF1B sg2 or HNF1B sg8 constructs, and carried out a subcutaneous tumour xenograft assay with the following experimental groups for each cell line: no dox from Day 0, +dox from Day 0, and +dox at a later time point once subcutaneous tumours had been established. This would have allowed me to compare the effects of knocking out HNF1B at the point of tumour cell seeding (i.e. +dox from Day 0) versus knocking out HNF1B in established tumours (i.e. +dox at a later time point, once tumours had already formed).

Control tumours (i.e. no dox from Day 0, expressing wild-type HNF1B) would be expected to grow rapidly, whereas tumours with HNF1B knockout from Day 0 would be expected to exhibit greatly reduced growth as seen previously. Meanwhile, knocking out HNF1B in established tumours might lead to tumour regression or simply a stagnation of tumour growth, or might have no effect at that late stage of tumour development. I might thus have been able to determine whether HNF1B was only important for tumour initiation, or whether it was also important for tumour maintenance.

Alternatively, for a more elegant and flexible *in vivo* experiment to test the importance of HNF1B in established tumours, I could have used HNF1B knockout clonal cell lines that lacked endogenous HNF1B but expressed the aforementioned untargetable HNF1B rescue construct, this time under the control of a dox-inducible promoter. This would essentially provide a conditional HNF1B knockout/restoration system, facilitated by dox withdrawal and dox

treatment respectively, to investigate the stage(s) of tumour development in which HNF1B plays a role.

To avoid having to transduce 786-M1A-TGL and OS-LM1B-TGL parental ccRCC cells with three different transgenes (i.e. Cas9 and HNF1B sg8, in addition to the dox-inducible HNF1B rescue construct) and their various antibiotic / fluorescent selection markers, I would generate the endogenous HNF1B knockout clones by transiently transfecting the cells (via electroporation) with a single plasmid, pX330[127], expressing both Cas9 and either the control sgRNA NTC18 or HNF1B sg8 (hereafter referred to as pX330-Cas9-NTC18 and pX330-Cas9-sg8 respectively). I would then sort for single cell clones 48hrs after electroporation, expand the clonal cell lines, and then screen them by HNF1B Western blot to identify endogenous HNF1B knockout clones.

One approach would be to first transduce 786-M1A-TGL and OS-LM1B-TGL parental cells with either the dox-inducible exogenous HNF1B construct or empty vector, and then electroporate those 786-M1A-TGL-tet-on-HNF1B and OS-LM1B-TGL-tet-on-HNF1B cells and select for single cell clones, all while maintaining the cells in +dox media. Alternatively, to avoid comparing two different clones in subsequent *in vivo* experiments, I could electroporate 786-M1A-TGL and OS-LM1B-TGL parental cells, select for single cell clones, and then transduce those clonal cell lines with either the dox-inducible exogenous HNF1B construct or empty vector. However, due to the strong selective disadvantage imposed by HNF1B knockout *in vitro* (***Figs 14 to 19***), the latter approach could have a higher risk of failing to yield HNF1B knockout clones and instead selecting for single cell clones that had either escaped knockout or adapted to HNF1B loss and become HNF1B-independent, prior to transducing the cells with dox-inducible exogenous HNF1B.

Ultimately, if I did manage to obtain clones that were endogenous HNF1B knockout but remained dependent on HNF1B (and which reliably expressed sufficient levels of exogenous HNF1B in a dox-inducible manner), along with suitable control clonal cell lines, I could have proceeded to carry out *in vivo* subcutaneous tumour xenograft assays to investigate whether HNF1B is important for tumour maintenance in established tumours, or whether it is only essential at the point of tumour initiation.

I would first have performed a pilot experiment *in vitro*, and carried out a proliferation assay with the following experimental groups: (i) +dox from Day 0, for the duration of the assay

(i.e. HNF1B restored from the point of cell seeding); (ii) no dox from Day 0, for the duration of the assay (i.e. HNF1B knockout from the point of cell seeding); (iii) +dox from Day 0 initially, followed by dox withdrawal at a later time point partway through the assay (i.e. simulating HNF1B knockout in established tumours).

Dox treatment would induce exogenous HNF1B expression, whereas dox withdrawal would remove exogenous HNF1B, effectively rendering the cells HNF1B knockout (as they already lack endogenous HNF1B protein). I would thus be able to observe whether dox withdrawal (and concomitant HNF1B depletion) at a later time point in the proliferation assay would result in decreased cell proliferation.

If the *in vitro* pilot experiment yielded positive results, I would then have used the clonal cell lines for subcutaneous tumour xenograft assays to investigate the effects of HNF1B depletion in established tumours, with 3 experimental groups: (i) +dox from Day 0 (i.e. HNF1B restored from the point of tumour initiation, with dox food pellets); (ii) no dox from Day 0 (i.e. HNF1B knockout from the point of tumour initiation, with normal food); (iii) +dox from Day 0 until tumours formed, at which point I would withdraw dox (by switching to normal food pellets). I would thus be able to observe whether depleting HNF1B in established tumours would lead to tumour regression (or perhaps simply a deceleration or cessation of tumour growth), or whether HNF1B loss has no effect at this later stage of ccRCC development. As additional controls, I would also incorporate additional dox treated and untreated experimental groups with the clonal cell lines that expressed the empty vector and were electroporated with pX330-Cas9-NTC18.

Finally, if I had had more time, it would also have been interesting to investigate whether HNF1B loss affected ccRCC metastasis. I chose to perform subcutaneous tumour xenograft assays for my project in order to investigate the importance of HNF1B for primary tumour initiation and growth, as opposed to *secondary* tumour initiation and growth. The process of metastasis involves additional "bottlenecks" to select for those tumour sub-clones capable of invading into the surrounding stroma, disseminating from the primary tumour mass to spread to distant sites (which requires intravasation, survival in the circulation, and then extravasation), and then surviving in and colonising the new tissue environments at these distant sites to form metastases[139]. Therefore, for my initial *in vivo* characterisation of HNF1B's role in ccRCC, it seemed logical to first focus on primary tumour initiation and growth, and the associated

bottlenecks at that earlier stage of tumour development, as opposed to the later stages of tumour progression when tumours become capable of forming metastases.

However, ccRCC can be highly aggressive: metastasis occurs in roughly a third of ccRCC cases, typically resulting in fatality[2–7]. Hence, when assessing HNF1B as a potentially targetable tumour dependency, its value as a therapeutic target would certainly increase if it was a critical transcription factor in both the primary tumour and secondary lesions, as this would suggest that targeting HNF1B could also help treat advanced metastatic disease.

One approach to study the importance of HNF1B for ccRCC metastasis would have been to inject 786-M1A-Cas9 and OS-LM1B-Cas9 Control and HNF1B knockout ccRCC cell lines intravenously (via the tail vein) into Nod-Scid-γ (NSG) immunocompromised mice. Owing to the high metastatic propensity of the 786-M1A-TGL and OS-LM1B-TGL parental cell lines[4], the cells should then disseminate to the lungs and seed lung metastases.

Subsequent metastatic tumour growth would then be measured weekly by bioluminescence imaging after intraperitoneal injection of D-luciferin[4]. I would thus be able to determine whether the metastatic potential of HNF1B knockout cells was reduced relative to wild-type Control cells, which would be indicated by the formation of fewer or smaller metastatic nodules in the lung.

I could also have performed additional experiments to study the importance of HNF1B in established metastases. For this, I could use the conditional HNF1B on/off system described above (i.e. endogenous HNF1B knockout clones expressing a dox-inducible exogenous HNF1B rescue transgene). I could then perform an experiment analogous to the aforementioned subcutaneous tumour xenograft assay designed to investigate the importance of HNF1B in established subcutaneous tumours, but with intravenous injection of the clonal cell lines. I would thus be able to observe whether dox withdrawal and concomitant HNF1B loss, occurring after metastases were established, would eliminate, shrink, or halt the growth of secondary lesions.

6) Identifying HNF1B transcriptional targets that might play a role in its function in ccRCC

After establishing the phenotype for HNF1B loss *in vitro* and *in vivo*, I started to explore the mechanisms underlying the selective disadvantage associated with HNF1B depletion in order

to elucidate the critical functions that HNF1B may perform in ccRCC. To that end, I carried out RNA-seq experiments with 786-M1A-Cas9 and OS-LM1B-Cas9 knockout (sg2 and sg8) and control (NTC18) cell lines. Meanwhile, a postdoc in the lab helped to carry out HNF1B ChIP-seq in several ccRCC cell lines, specifically two parental ccRCC cell lines, 786-O and OS-RC2, and their more metastatic derivatives (786-M1A and OS-LM1B respectively). These experiments helped to identify downstream transcriptional targets of HNF1B that may be directly regulated by this transcription factor, and which may contribute to its critical role in ccRCC.

The RNA-seq data revealed, among other transcriptional changes, that MYC and BCL2 were downregulated in HNF1B knockout cells relative to control cells (***Fig 27A-B, E***). Consistent with this, gene set enrichment analysis (GSEA) of the RNA-seq data revealed an enrichment of MYC target genes in control versus knockout cells (***Fig 28B***). These results were consistent between both the 786-M1A-Cas9 and OS-LM1B-Cas9 cell line systems.

Meanwhile, the HNF1B ChIP-seq revealed multiple HNF1B binding peaks in the *MYC* and *BCL2* loci (***Fig 29A-B***). These HNF1B binding peaks were shared between all four ccRCC cell lines (i.e. both parental ccRCC cell lines, 786-O and OS-RC2, and both of their metastatic derivatives, 786-M1A and OS-LM1B). Moreover, the HNF1B binding peaks near the *MYC* and *BCL2* genes contained transcription factor motifs that bore a striking resemblance to the known HNF1B transcription factor motif.

Altogether, the data from the RNA-seq and HNF1B ChIP-seq (plus HNF1B transcription factor motif analysis) experiments indicated that MYC and BCL2 are downregulated upon HNF1B loss, and could possibly be direct transcriptional targets of HNF1B. In light of these results, it would also have been interesting to check whether there are similar changes in MYC and BCL2 expression in patient tumour cohorts.

Furthermore, it would have been interesting to combine the RNA-seq and HNF1B ChIP-seq data and perform a more comprehensive analysis to determine whether the other genes that are differentially expressed in the RNA-seq data (apart from the aforementioned EMT genes that were upregulated, and the *MYC*, *BCL2*, and *CDH16* genes that were downregulated, in HNF1B knockout cells) are also bound by HNF1B according to the HNF1B ChIP-seq data.

In addition, some of the HNF1B binding peaks near the *MYC* locus also spanned the genomic region around *PVT1* (plasmacytoma variant translocation 1; ***Fig 29A***), which produces a long non-coding RNA (lncRNA), PVT1. Increased expression of *PVT1* and the PVT1 lncRNA

has been implicated in tumourigenesis, tumour progression, and poor prognosis in multiple cancer types[140–148]. Moreover, the PVT1 lncRNA reportedly promotes MYC expression by protecting MYC protein from degradation[140–148]. Increased PVT1 expression and concomitant MYC upregulation have apparently been observed and implicated in several human cancers, including in ccRCC[140–148]. One study analysing TCGA data suggested that ccRCC (denoted "KIRC", i.e. kidney renal clear cell carcinoma, in the TCGA database) showed particularly upregulated *PVT1* expression compared to other tumour types, and that this was associated with high MYC protein levels and subsequent deregulation of MYC target genes, all of which correlated with poorer prognosis[144].

On the other hand, there is another study which suggests that the *PVT1* promoter regulates *MYC* independently of the PVT1 lncRNA and actually inhibits *MYC* transcription[140,143]. This study concluded that *PVT1*-mediated regulation of *MYC* ultimately maintains MYC levels within a limited range, with the *PVT1* promoter exerting a more dominant regulatory effect and inhibiting *MYC* transcription, while the PVT1 lncRNA stabilizes MYC protein and thus promotes *MYC* expression at the protein level [140,143]. Therefore, there is disparate evidence in the literature regarding the precise regulatory effects of *PVT1* on *MYC* expression. There are also studies suggesting that MYC, in turn, regulates the expression of *PVT1*[141,142,145,147,148].

It might thus have been interesting to investigate the potential involvement of *PVT1* (its promoter and/or the PVT1 lncRNA) in HNF1B regulation of MYC, and in the phenotypes associated with HNF1B loss and subsequent MYC downregulation (such as the reduced proliferation and tumour growth, defect in cell cycle progression, and increased apoptosis observed upon HNF1B knockout).

I next validated the results of the RNA-seq with qPCR to confirm the downregulation of these genes following HNF1B knockout. I used independent RNA samples taken from 786-M1A-Cas and OS-LM1B-Cas9 Control, sg2, and sg8 cell lines. These RNA samples were extracted from cell pellets taken during the competition proliferation assays performed with these two sets of cell lines (***Fig 14A-B***).

The MYC and BCL2 qPCRs performed with the 786-M1A-Cas9 and OS-LM1B-Cas9 RNA samples also indicated a decrease in MYC and BCL2 expression in sg2- and sg8-expressing cell lines relative to the Control cell lines, and were thus consistent with the RNA-seq data showing a downregulation of MYC and BCL2 in knockout cells relative to controls for the

same sets of cell lines (***Fig 30A-D, G***). However, the magnitude of the decreases in MYC and BCL2 expression observed by qPCR exceeded the downregulation observed in the RNA-seq data (***Fig 30A-D, G***).

The sets of 786-M1A-Cas9 and OS-LM1B-Cas9 Control, sg2 and sg8 cell lines used for the RNA-seq were generated from an independent lentiviral transduction, separate from the one used to generate the 786-M1A-Cas9 and OS-LM1B-Cas9 Control, sg2 and sg8 cell lines for the competition assays. Therefore, there may have been differences in the transduction efficiency of the sg2 and sg8 constructs that could explain the different magnitudes of gene expression changes between the two experiments. Alternatively, there may have been technical differences in the RNA extraction or reverse transcription of the RNA samples used for the qPCR validation, and/or variation in the amounts of cDNA used in the qPCR for these samples, which could have led to discrepancies in the observed changes in MYC and BCL2 mRNA levels.

Nevertheless, despite the different magnitudes of reduction in MYC and BCL2 expression, the qPCR data confirmed the conclusion from the RNA-seq data that MYC and BCL2 are both downregulated following HNF1B loss.

7) The phenotype for HNF1B loss may be due to defective cell cycling and increased cell death

To continue characterising the mechanisms underlying the selective disadvantage associated with HNF1B depletion (and thus the mechanisms by which HNF1B might support ccRCC), I next performed experiments to investigate the nature of the phenotype associated with HNF1B knockout, in order to determine which cellular processes were responsible for the selective disadvantage observed.

I hypothesised that the selective disadvantage conferred by HNF1B depletion could be due to increased cell death, or impaired progression through the cell cycle and consequently decreased cell proliferation, or possibly a combination of both mechanisms. I thus carried out assays to check for increased apoptosis and defects in cell cycle progression in HNF1B knockout cells.

Unfortunately, I did not have time to optimise and carry out experiments to investigate the possible contributions of other phenotypes (such as cellular senescence or differentiation) to

the selective disadvantage associated with HNF1B knockout. If I had had more time on this project, I would also test whether other mechanisms, such as increased senescence, also played a role in the loss of ccRCC cell fitness after HNF1B depletion.

The cell cycle assay data indicated a defect in the G1/S transition in HNF1B knockout cells, resulting in delayed progression through the cell cycle relative to Controls (***Fig 31, Fig 32***). This defect was rescued by the expression of exogenous HNF1B (***Fig 32***). Meanwhile, the Annexin V / Propidium iodide staining for apoptosis revealed a substantial increase in cell death (specifically in the percentage of late apoptotic / necrotic / dead cells) in HNF1B knockout cells relative to Controls, which was also rescued by exogenous HNF1B expression (***Fig 33***).

I carried out the cell cycle assays and Annexin V / Propidium Iodide staining on 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cells, as well as 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells.

Ideally, I would have tested for increased apoptosis and cell cycle defects in CRISPRi cells expressing HNF1B CRISPRi #2 compared to the Scramble Control, to see if the results obtained with the CRISPR/Cas9 cell lines were still the same with CRISPRi cell lines, in which HNF1B is depleted without the formation of double-stranded breaks (DSBs) in the DNA. This could have been informative, as DSB formation might affect apoptosis and cell cycling. However, these CRISPRi cell lines are mCherry positive, and were therefore not suitable for use in Annexin V and/or PI staining experiments, due to the incompatibility with PI.

If I had had more time I would have cloned BFP versions of the CRISPRi cell lines in order to carry out cell cycle assays and Annexin V / PI staining with this gene knockdown system as well. In addition, I would also have repeated the cell cycle and Annexin V / PI experiments, both with freshly transduced versions of the same 786-M1A-Cas9 and OS-LM1B-Cas9 Control-BFP, sg2-BFP, sg8-BFP, Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cell lines, and also with other ccRCC cell line systems.

The cell cycle defect in HNF1B knockout cells at the G1/S transition was perhaps most clearly evident in the cell cycle assay performed with the 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines (***Fig 31***). The cell cycle assay with the 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells also strongly supported the conclusion that HNF1B knockout cells experience a defect in transitioning from G1 to S phase (***Fig 32***). However, as these cell lines were transduced independently and differently (i.e.

they also contain either empty vector or the HNF1B rescue construct, in addition to the sgRNAs), their cell cycle distributions varied somewhat compared to their 786-M1A-Cas9 Control-BFP and sg8-BFP counterparts.

Despite this, the defect was still noticeable in both experiments. Moreover, evidence of the defect emerged at the same time point (t12) and became increasingly apparent over the same time period (t12-t16), in each time course (***Fig 31, Fig 32***). If I had had more time, I would have carried out additional experiments to dissect the possible mechanisms by which HNF1B knockout might cause this impairment in cell cycle progression, at a genetic and molecular level.

In both Annexin V / PI staining experiments (the first performed with 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cells (***Fig 33A-C***), and the second performed with 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cells (***Fig 33D-F***)), there was a small amount of cell death detected in the control cell lines (***Fig 33A, D***). The presence of late apoptotic / necrotic / dead cells in the control cell lines in each experiment could be the result of technical issues with sample processing, which can lead to cell damage and thus a positive Annexin V and PI signal (for instance, the use of Trypsin to detach apoptotic cells, rather than a gentler agent such as Accutase). Therefore, if I had had more time on this project, I would optimize and repeat the Annexin V / PI staining experiments to see if I could remove or at least reduce this background level of cell death.

Moreover, in both Annexin V / PI staining experiments, the percentages of Annexin V+/PI– (i.e. early apoptotic) cells remained similar between the different cell lines (***Fig 33***). Hence, both experiments suggested that HNF1B loss does result in increased levels of cell death compared to both HNF1B wild-type control cells and HNF1B-restored cells, and that this difference mainly lies in the proportion of late apoptotic / necrotic / dead cells.

One possible explanation as to why I only observed differences in the percentage of late apoptotic / necrotic / dead cells is that cells in early apoptosis may have been further damaged and thus progressed to late stage apoptosis during the processing and staining of the cell samples, thus resulting in the low percentages of early apoptotic cells detected in each assay and the lack of changes in early apoptosis between cell lines.

Alternatively, the predominance of late apoptotic / necrotic / dead cells and the relative lack of early apoptotic cells could be due to the fact that the cell lines used were not freshly transduced; they were thawed from cell stocks and then expanded in culture for some time before

use. It is possible that the cells may have progressed from early apoptosis to late stage apoptosis, necrosis, or outright cell death during this time period in culture. This may be especially likely for the sg2, sg8, and sg8(+EV) knockout cell lines, where even a limited time expanding in culture would mean a prolonged length of time experiencing the effects of HNF1B loss. Consequently, the apoptotic phenotype may have increased in severity over time in culture before I stained the cells with Annexin V and PI. Thus, although a higher percentage of early apoptotic cells may have been present at some stage after sgRNA transduction in the control, knockout, and rescue cell lines, I might have missed the appropriate time window in which to observe these early apoptotic cells. Using freshly transduced cells might have prevented this issue.

On the other hand, each cell line originally underwent antibiotic selection after viral infection to select for successfully transduced cells. The cell death during antibiotic selection would have obscured any cell death resulting from HNF1B loss, and I was thus unable to measure apoptosis from the onset of HNF1B depletion after infection with sgRNA virus. Therefore, if HNF1B knockout did induce apoptosis, the early stages of apoptosis may have already elapsed by the time I froze down cell stocks, and I might already have missed the optimal time window to detect early apoptotic cells.

However, further repeats of these Annexin V / PI experiments may have revealed whether the differences in cell death between control / rescue and knockout cells consistently appeared in the percentages of late apoptotic / necrotic / dead cells, with no changes in early apoptosis. If so, this might suggest that this was in fact a real biological result. For instance, this could indicate that the intensity of the cell death stimulus triggered by HNF1B loss in ccRCC cells is so strong that HNF1B knockout cells either progress very rapidly through early apoptosis into late apoptosis, or undergo necrotic cell death instead. Consequently, I was mainly able to detect late apoptotic or necrotic cells with my Annexin V/PI staining, rather than early apoptotic cells.

Cells undergoing late apoptosis and cells undergoing necrosis both stain as Annexin V+/PI+, and thus cannot be distinguished by Annexin V / PI staining; rather, cells undergoing both types of cell death would appear in the same population. Necrosis can occur after apoptosis, or happen in parallel. Moreover, cells may enter necrosis instead of apoptosis in response to the same stimulus, if they experience that stimulus at a higher intensity or for a longer duration.

Ongoing apoptosis may even be converted to necrosis, for instance due to reduced availability of ATP or caspases inside the dying cell[149]. Therefore, while it is possible that the HNF1B knockout cells stained in the Annexin V / PI experiments were in late apoptosis, it is also possible that they had instead undergone necrotic cell death.

If I had had more time to further investigate this, I would have performed additional experiments to confirm whether HNF1B knockout cells were undergoing apoptosis, and if so, to determine the stage of apoptosis that HNF1B knockout cells progressed to. For example, I would have carried out caspase activity assays (such as the Promega Caspase-Glo 3/7 Assay) and Western blots for caspase-3 / cleaved caspase-3 *in vitro*, or TUNEL assays and caspase-3 IHC with tumour sections to check for apoptosis *in vivo*. In addition, I would have carried out experiments to test for necrosis, and distinguish it from apoptosis.

Based on the results of experiments such as these experiments, and on the results from repeats of the Annexin V / PI staining, I could then have carried out further experiments to charactserise the possible mechanisms by which HNF1B knockout might trigger the type of cell death in question at a genetic and molecular level.

Ultimately, despite the possible influence of technical issues while performing these Annexin V / PI staining assays, and despite the unknown reasons as to why differences in cell death seem to only only appear in the percentages of late apoptotic / necrotic / dead cells, the data still clearly showed a substantial increase in cell death in HNF1B knockout cells relative to control cells, which was completely reversed by HNF1B restoration (***Fig 33***). The technical and biological explanations speculated on above would be unlikely to detract from this conclusion, especially since the increase in cell death between knockout and control cells was consistent between both Annexin V / PI experiments (i.e. between both sets of HNF1B knockout versus isogenic control cell lines (***Fig 33A, C-E***)), suggesting that this is a real biological result.

Therefore, taken together, the data from the cell cycle PI and Annexin V / PI experiments suggested that the selective disadvantage observed with HNF1B loss *in vitro* and *in vivo* stemmed from a combination of decreased cell proliferation due to impaired cell cycle progression, and increased cell death. Thus, HNF1B's role in ccRCC might be to promote cell proliferation and cell survival.

8) Potential roles of MYC and BCL2 in HNF1B's function in ccRCC: BCL2 involvement

Combining this data with the RNA-seq, HNF1B ChIP-seq and HNF1B transcription factor motif analysis, I hypothesised that the decreased expression of the directly regulated HNF1B target genes MYC and BCL2 might be responsible for the selective disadvantage associated with HNF1B loss, and thus that MYC and BCL2 might be involved in HNF1B's role in ccRCC.

I thus decided to investigate the relative contributions of MYC and BCL2 to the selective disadvantage conferred by HNF1B depletion. I carried out CRISPRi competition proliferation assays, comparing the effects of knocking down MYC and BCL2 (individually or in combination) against the selective disadvantage conferred by HNF1B knockdown (***Fig 35A-B***).

I chose to use the CRISPRi system for these experiments instead of the CRISPR/Cas9 system in order to more accurately recapitulate the downregulation of MYC and BCL2 observed by RNA-seq in HNF1B knockout cells. I reasoned that CRISPRi knockdown would more closely recapitulate a downregulation of these genes, whereas the CRISPR/Cas9 system would be more likely to mediate a complete knockout of MYC and BCL2. A complete loss of MYC and BCL2 expression might induce too strong a phenotype that would not be physiologically relevant, since the expression of these genes is not completely lost upon HNF1B depletion.

I designed functional MYC and BCL2 CRISPRi tandem sgRNA constructs that mediated efficient knockdown of their target genes, as verified by qPCR (***Fig 34***). The HNF1B, MYC and BCL2 CRISPRi tandem sgRNA constructs used also mediated successful target gene knockdown during the competition assays, again verified by qPCR (***Fig 35C-H***).

Based on the data from these MYC/BCL2 CRISPRi competition proliferation assays, MYC knockdown phenocopied the selective disadvantage observed with HNF1B depletion, producing a similarly strong (or even stronger) phenotype, whereas BCL2 knockdown had no apparent effect (***Fig 35A-B***).

The lack of a phenotype with BCL2 knockdown was an unexpected result. One possibility is that BCL2 loss might not be the main or sole instigator of the enhanced cell death seen in HNF1B knockout cells. It is likely that other factors (which may or may not be regulated by HNF1B) are required to mediate cell death. Thus, BCL2 knockdown alone would be unable to recapitulate the phenotype seen with HNF1B knockdown.

Alternatively, since the increased cell death observed in knockout cells in the Annexin V / PI staining was in the late apoptotic / necrotic / dead cell population, it is possible that HNF1B knockout induces necrosis instead of apoptosis, which might not involve BCL2. Hence, BCL2 knockdown would not phenocopy the effects of HNF1B knockdown, although this does raise the question as to what impact BCL2 downregulation has in HNF1B-depleted ccRCC cells.

Another possibility is that the increased cell death seen in the HNF1B knockout cell lines could be an indirect effect of HNF1B depletion. I will proceed to describe this hypothesis only with reference to apoptotic cell death, although it may also be applicable to other types of cell death (such as necrosis), should those pathways be involved.

Instead of HNF1B loss directly affecting the expression of genes involved in apoptosis to trigger pro-apoptotic signals and cause increased cell death, HNF1B loss might only render ccRCC cells more sensitive to apoptotic stimuli. HNF1B knockout cells may subsequently undergo apoptosis in response to an additional event (such as MYC downregulation) caused by HNF1B depletion, resulting in the increased cell death detected by Annexin V/ PI staining.

The downregulation of BCL2 following HNF1B knockout may even play a role in facilitating this increased sensitivity to apoptotic signals, rather than directly inducing cell death by itself. BCL2 is an anti-apoptotic protein, and may thus serve to protect ccRCC cells from cell death when its expression is upregulated by wild-type HNF1B. When BCL2 is downregulated upon HNF1B knockout, this suppression of apoptosis would be removed, thus leaving the cells more sensitive to pro-apoptotic signals. Apoptosis might then be induced in response to an additional event directly precipitated by HNF1B knockout that generated a pro-apoptotic signal. For instance, the downregulation of MYC (and the consequent withdrawal of oncogenic signalling and impaired cell cycle progression) might generate apoptotic signals and thus lead to increased cell death in this context.

It is also possible that HNF1B loss renders ccRCC cells more susceptible to cell death through a mechanism independent of BCL2 downregulation, and that BCL2 downregulation could be the additional event that induced apoptotic signals and caused increased cell death (but only in the context of HNF1B depletion).

If this hypothesis is true, then perhaps BCL2 knockdown alone did not phenocopy the effects of HNF1B knockdown because BCL2 loss in itself is not sufficient to induce apoptosis and instead only renders ccRCC cells more sensitive to apoptotic signals. However, there was no

subsequent genetic event to induce cell death in these BCL2 knockdown cells, especially since BCL2 was knocked down in otherwise HNF1B wild-type, MYC wild-type ccRCC cells. Therefore, if BCL2 knockdown does confer increased cell sensitivity to apoptosis and MYC withdrawal is the key death-triggering stimulus required, the cells would become vulnerable to apoptosis but would not exhibit a phenotype due to the presence of wild-type MYC. If instead the sensitivity to apoptosis is conferred by a BCL2-independent mechanism stemming from HNF1B depletion, then these cells would lack an increased sensitivity to apoptosis due to the presence of wild-type HNF1B, and thus might not readily respond to any pro-apoptotic signals induced by BCL2 loss.

Ultimately, this hypothesis might reconcile the observations of increased cell death in HNF1B knockout cells, and yet the lack of a phenotype with BCL2 knockdown alone in cells that were otherwise HNF1B wild-type and MYC wild-type. It would also suggest that perhaps the knockdown of another HNF1B target gene in those BCL2-depleted cells (like MYC, for instance) may have triggered apoptosis and caused the same striking phenotype seen with HNF1B knockdown and MYC knockdown alone. In support of this hypothesis, there is some evidence in the literature suggesting that the downregulation or inhibition of BCL2 can cooperate to enhance apoptosis induced by drug treatment[150,151].

If I had had more time on this project, I would have performed cell cycle and apoptosis assays with freshly transduced versions of the cell lines used for the MYC/BCL2 CRISPRi competition proliferation assays (***Fig 35A-B***), in order to determine whether MYC and BCL2 knockdown (individually and in combination) also resulted in a cell cycle defect in G1/S progression and increased cell death, similar to that observed in HNF1B knockout cells.

Ultimately, increased cell death might still contribute to the selective disadvantage seen with HNF1B knockout, along with impaired cell cycling and reduced cell proliferation. However, this increased cell death may not be triggered directly by BCL2 downregulation, but might instead be induced by the downregulation of MYC following HNF1B loss.

9) Potential roles of MYC and BCL2 in HNF1B's function in ccRCC: MYC involvement

Deregulated MYC expression and/or activation is implicated in a variety of different cancer types[130–132]. Among its myriad of functions, MYC regulates progression through various

stages of the cell cycle by regulating the expression of key cell cycle proteins. Intriguingly, there is considerable evidence suggesting that MYC promotes progression from G1 to S phase.

For example, MYC upregulates various cyclin and cyclin-dependent kinase (Cdk) genes involved in G1 phase; regulates E2F transcription factors (which are essential for G1/S progression, as they in turn regulate the expression of cell cycle proteins required for S phase), and represses the expression and activity of CIP/KIP proteins (cyclin-dependent kinase (Cdk) inhibitors that can inhibit G1/S progression).[132]

Moreover, various studies have demonstrated that modulating MYC expression and activity has effects on the G1/S transition in different cell line and tumour systems. For example, MYC overexpression can stimulate quiescent cells to enter the cell cycle[132] and drive cells to enter S phase[130,132].

By contrast, depletion or inactivation of MYC, or inhibition of MYC function, has been shown to prevent quiescent cells from re-entering the cell cycle, hinder cell cycle progression, prevent entry into S phase, and cause cell cycle arrest at G0/G1[132,152]. Expression of a dominant negative mutant of MYC, for instance, causes cell cycle arrest at G1, which can be rescued by wild-type MYC expression[132]. Similarly, MYC-deficient cells exhibit delayed exit from the G1 phase of the cell cycle, and also exhibit slower growth that can be rescued by MYC expression[132].

Unsurprisingly, evidence in the literature indicates that high levels of MYC expression and activity enhances cell proliferation and tumour growth[130–132], whereas MYC downregulation or functional antagonism incurs reduced proliferation and tumour regression (typically due to the onset of proliferative arrest, apoptosis, senescence, and/or cancer cell differentiation).

It is therefore encouraging to note that many of the phenotypes identified upon MYC withdrawal in the literature were also reflected in my own data in HNF1B knockout cells *in vitro* and *in vivo*, and in MYC knockdown cells *in vitro*. Following HNF1B knockout and concomitant downregulation of MYC *in vitro*, I also observed decreased cell proliferation, increased cell death, and impaired cell cycle progression, specifically a defect at the G1/S transition (which resulted in a delay in exiting G1 and entering S phase, and thus a prolonged G1 phase), in HNF1B knockout cells compared to HNF1B wild-type controls. These phenotypes were all reversed by HNF1B restoration, and presumably associated MYC restoration, *in vitro*. Moreover,

I observed reduced tumour growth with HNF1B knockout *in vivo*, and a similarly striking reduction in cell proliferation with direct MYC knockdown by CRISPRi *in vitro*.

If I had had more time, I would also have explored other phenotypes in the aftermath of HNF1B knockout and MYC downregulation (such as senescence or tumour cell differentiation), and investigated the contributions from other MYC functions (such as its ability to promote ribosome biogenesis and alter cell metabolism, which would help fuel cell growth to accompany the increased proliferation driven by MYC).

There is substantial evidence in the literature indicating that MYC activation by itself is insufficient to promote tumourigenesis, as it is countered by mechanisms such as proliferative arrest, senescence, and/or apoptosis[130–132]. Hence, MYC requires a permissive genetic and/or epigenetic context, and the cooperation of other oncogenes or genetic alterations that impede these cellular counter-measures, in order to facilitate tumourigenesis[130–132]. MYC expression or activity in tumours is thus frequently combined with the acquisition of additional oncogenic mutations, the loss of tumour suppressor genes, the loss of senescence regulators, and/or the increased or decreased expression of anti-apoptotic or pro-apoptotic regulators respectively.

In the case of ccRCC, tumour cells have already undergone the loss of a tumour suppressor, VHL, resulting in the accumulation and deregulated transcriptional activity of HIF2A. Against this background of reduced tumour suppression and aberrant HIF signalling, HNF1B upregulates the anti-apoptotic protein BCL2. This increased BCL2 expression might thus protect ccRCC cells against apoptosis triggered by elevated MYC expression, which is also induced by HNF1B.

Of course, there may also be additional gene regulatory events either downstream of the VHL-HIF pathway (as a result of VHL loss and/or HIF2A transcriptional activity) or downstream of HNF1B (due to its own transcriptional activity), as well as other genetic/epigenetic alterations and their consequences in ccRCC, that may cooperate with MYC to overcome cellular checkpoint mechanisms and drive tumourigenesis and tumour progression. However, HNF1B's upregulation of both MYC and BCL2 (thereby mediating both enhanced cell proliferation and increased resistance to apoptosis) on a background of deregulated VHL-HIF signalling (which is already known to be an important driver event in ccRCC tumourigenesis) is one possible mechanism underlying tumour formation and progression.

Several studies suggest that there is a threshold level of MYC expression that must be maintained in order to support tumourigenesis[130,131]. Evidence in the literature also suggests that MYC depletion can lead to apoptosis in tumour-derived cell lines *in vitro* or in established tumours *in vivo*[130–132]. Moreover, various studies in different tumour types have also shown that MYC depletion or inhibition in established tumours leads to tumour regression, suggesting that many tumours become addicted to MYC and are thus dependent on continuous MYC expression and activity for tumour maintenance[130,131].

In line with this, it is possible that the increased cell death observed in HNF1B knockout cells in the Annexin V / PI assays could in fact be caused by the MYC downregulation associated with HNF1B loss. If so, this might suggest that ccRCC cells are addicted to MYC, and might also reconcile the data showing enhanced cell death in HNF1B knockout cells, and yet the absence of a phenotype with BCL2 knockdown by CRISPRi.

If the upregulation of BCL2 in HNF1B wild-type cells does help to prevent apoptosis in response to MYC overexpression, then the downregulation of BCL2 upon HNF1B knockout might leave ccRCC cells more sensitive to apoptotic signals. The concurrent downregulation of MYC following HNF1B loss might then induce apoptosis. Of course, other gene expression changes downstream of HNF1B loss could also be involved, apart from the downregulation of these two genes. However, the increased cell death seen in HNF1B knockout cells could be due, at least in part, to the combined loss of MYC and BCL2, with reduced BCL2 expression sensitizing the cells to pro-apoptotic signals (but not eliciting a phenotype on its own) and MYC withdrawal triggering cell death.

10) HNF1B may be important in ccRCC due to its ability to regulate MYC

Ultimately, it is possible that ccRCC cells are dependent on HNF1B due to its ability to regulate MYC and facilitate MYC-driven tumourigenesis. HNF1B may drive MYC overexpression and maintain high levels of MYC, above the threshold required to promote ccRCC cell proliferation, tumour formation and tumour growth. Additionally, HNF1B transcriptional activity may help provide a permissive genetic and/or epigenetic context, on a background of VHL tumour suppressor loss and high HIF2A activity, for MYC to promote

ccRCC tumourigenesis (for example, HNF1B upregulates the anti-apoptotic protein BCL2, which might prevent apoptosis in response to high MYC expression and activity).

However, the hypothesis that HNF1B is important in ccRCC *in vitro* and *in vivo* primarily due to its ability to upregulate MYC still needs to be confirmed definitively. For example, if I had had more time, I would have performed competition assays with MYC restoration in HNF1B knockout cells to determine whether exogenous MYC expression could rescue the selective disadvantage, defective G1/S progression, and increased cell death associated with HNF1B loss, to a similar extent as the rescue conferred by exogenous HNF1B expression (***Fig 23A-D***). I would also have investigated whether MYC CRISPRi knockdown cells, which exhibited a similar selective disadvantage to HNF1B CRISPRi knockdown cells (***Fig 35A-B***), also showed increased cell death and a similar cell cycle defect at the G1/S transition.

Evidence in the literature suggests that MYC is important for both tumour initiation and tumour maintenance[130,131]. However, in the context of my project, is uncertain as to whether or not HNF1B (and MYC) are important for both tumour initiation and maintenance in ccRCC. It is thus also uncertain whether ccRCC cells are addicted to HNF1B and/or MYC. Further study would be required to address these questions, including carrying out *in vivo* experiments to investigate whether HNF1B (and MYC) are essential for tumour maintenance, as described earlier in this section.

Altogether, the reduced proliferation and tumour growth associated with HNF1B knockout *in vitro* and *in vivo*, along with the enhanced cell death observed in HNF1B knockout cells *in vitro* and the possibility that this increased apoptosis may be linked to MYC downregulation, do seem to suggest that ccRCC cells may be addicted to HNF1B and possibly also to MYC. It would thus be important to investigate whether HNF1B depletion or direct MYC depletion in established tumours would lead to tumour regression. If so, this would support the hypothesis that ccRCC cells are addicted to HNF1B and/or MYC.

Further study would also be required to investigate the relationship between the HNF1B-MYC axis and the well-characterised VHL-HIF2A axis (already known to be essential driver pathway in ccRCC). There is evidence in the literature to suggest that HIF2A enhances MYC activity in VHL-deficient ccRCC, and that the two transcription factors may cooperate to mediate tumour progression[153,154].

11) Conclusions and implications

Characterising the mechanisms by which HNF1B supports ccRCC, and the role of MYC in these mechanisms, could have implications for both the study of transcription factors in cancer, and ccRCC therapy.

My data indicates the possibility that ccRCC cells depend on HNF1B due to its regulation of MYC. MYC is a commonly deregulated oncogenic driver in a diverse range of tumour types[130–132]. By contrast, HNF1B seems to be important in only a limited range of tumour types, including ccRCC[155,156], and is a renal developmental transcription factor that is highly expressed in ccRCC compared to other tumour types, but is not subject to genetic alterations in ccRCC. There is a possibility that HNF1B may belong to a renal-specific core transcriptional network, along with other essential ccRCC transcription factors like HIF2A, which converges on the regulation of common oncogenes (such as MYC) or oncogenic pathways in order to support malignant phenotypes and thereby promote ccRCC tumourigenesis and tumour progression.

This is a potentially interesting notion if it turns out to be generally applicable to other lineage transcription factors that also play essential roles in cancer, since lineage transcription factors often exhibit relatively tumour-type specific importance[155,156]. Perhaps lineage transcription factors so often represent tumour dependencies because they also belong to tissue-specific core transcriptional networks that drive common oncogenic events, and thus the same oncogenes are regulated by different transcriptional pathways in different tumours derived from different tissue types.

My data also highlights the potential of HNF1B, and/or its transcriptional target MYC, as therapeutic targets in ccRCC treatment. Although both HNF1B and MYC are transcription factors and may thus be considered “undruggable”, there is already considerable evidence in the literature for the utility and feasibility of targeting transcription factors in cancer therapy[86,95]. As detailed above and in the Introduction section, examples of such transcription factor targets include: nuclear steroid receptors in breast and prostate cancer, the transcription factor MITF in drug-resistant BRAF-mutant melanoma[96], the oncogenic transcription factor MYC in multiple tumour types, and HIF2α in ccRCC[75,76].

Promising steps have already been made towards targeting MYC in other tumour types[130–132], and this might be an effective means of both inhibiting MYC function in ccRCC and also targeting HNF1B through one of its target genes. Alternatively, similar strategies to those used to target MYC and HIF2A may be applied to target HNF1B. Indeed, HNF1B may be the more attractive target as it lies upstream of, and regulates, MYC.

Ultimately, both HNF1B and MYC seem to be essential ccRCC tumour dependencies, and may thus be useful therapeutic targets in ccRCC, either individually or in combination with other therapeutic strategies to try to prevent potential issues with the development of tumour drug resistance.

12) Summary

In summary, I performed an *in vitro* CRISPR/Cas9-based genetic depletion screen with an sgRNA library targeting 50 transcription factors that are highly expressed in ccRCC. This screen highlighted the renal developmental transcription factor HNF1B as a potential ccRCC cell dependency. I thus proceeded to validate the importance of HNF1B in ccRCC, and to investigate the mechanisms underlying its role in ccRCC cells.

My data demonstrated a striking selective disadvantage with HNF1B loss *in vitro* and *in vivo*, which was consistent with the significant depletion of HNF1B sgRNAs in the screen, and suggested that this highly expressed transcription factor plays an important role in ccRCC. My data also indicated that HNF1B knockout resulted in a defect in cell cycle progression at the G1/S transition, as well as increased cell death (two mechanisms that might contribute to the selective disadvantage associated with HNF1B loss). Additionally, my data highlighted the HNF1B transcriptional targets MYC (which regulates cell cycle progression) and BCL2 (which functions as an anti-apoptotic protein) as target genes that might be directly regulated by HNF1B, and which might be at least partly responsible for these phenotypes (since both seemed to be downregulated upon HNF1B knockout). I subsequently showed that individual knockdown of MYC phenocopied HNF1B loss, conferring a striking selective disadvantage in ccRCC cells similar to that seen with HNF1B depletion. By contrast, individual knockdown of BCL2 had no observable effect.

Taken together, my data suggested that HNF1B-mediated regulation of MYC expression might be a key mechanism underlying the importance of HNF1B in ccRCC. HNF1B's ability to upregulate MYC might help activate MYC-driven transcriptional programmes that promote and support ccRCC cell proliferation and tumour growth. When HNF1B is depleted, MYC downregulation might result in impaired G1/S progression, and this cell cycle defect might subsequently inhibit proliferation. Moreover, MYC withdrawal might trigger increased cell death.

These possibilities seem to be consistent with evidence in the literature from other tumour types with high MYC expression: such tumours become dependent on MYC-driven growth programmes, and thus become addicted to MYC; upon MYC withdrawal, these MYC-addicted tumours often exhibit cell death and regress, at least temporarily.

Although BCL2 knockdown alone did not show a phenotype, HNF1B-mediated regulation of BCL2 might still contribute by conferring increased resistance to apoptosis in HNF1B wild-type cells (since apoptosis is one of several cellular responses to counter MYC activation). BCL2 downregulation in HNF1B-depleted cells might thus confer increased susceptibility to apoptosis triggered by MYC withdrawal.

To conclude, based on my data I would tentatively propose that one possible mechanism by which HNF1B supports ccRCC is through upregulating MYC, and that ccRCC tumours may in fact be addicted to MYC and therefore to HNF1B, explaining their dependency on this highly expressed, non-mutated renal developmental transcription factor.

My data adds to the growing body of evidence in the literature implicating other lineage transcription factors in different tumours types. My results would also encourage further study and characterisation of the "HNF1B-MYC axis" in these tumours, as well as any relationship between HNF1B-MYC and the VHL-HIF2A pathway already known to drive tumourigenesis in ccRCC. Finally, my results also highlight HNF1B and MYC as potentially useful therapeutic targets in ccRCC, perhaps in combination with other therapeutic strategies.

MATERIALS & METHODS

In vitro cell culture

293T cells (used for lentivirus production), HeLas, and MDA-MB-231 breast cancer cells were cultured in high glucose DMEM 1X (Dulbecco's Modified Eagle Medium, containing 4.5g/L D-glucose, L-glutamine and pyruvate; *Life Technologies Gibco, cat. no. 41966-029* and *41966-052*), supplemented with 10% foetal calf serum (FCS) & 1% pen-strep (P/S). HCC-1419 breast cancer cells, ccRCC cell lines and their derivatives were cultured in RPMI-1640 medium (*Sigma, cat. no. R8758*) supplemented with 10% FCS and 1% P/S. For dox treatment, doxycycline (*Sigma, cat. no. D9891*) was diluted in cell culture media from a stock concentration of 1mg/ml to the working concentration of 1µg/ml. The dox was initially added to the cells during cell seeding, and subsequently replenished with media changes and the accompanying addition of fresh dox every 2-3 days.

Lentiviral transduction

293T cells were used for lentivirus production. psPAX2 (*Addgene plasmid #12260*) and pMD2.G (*Addgene plasmid #12259*) were used for viral packaging plasmids. 293Ts were seeded at a density of $1x10^6$ cells per well in 6-well plates, with a total volume of 2ml media per well, and allowed to adhere overnight. 293Ts were then transfected with either FuGENE (*FuGENE 6 Transfection Reagent, Promega, cat. no. E2693*) or Lipofectamine 2000 (*Lipofectamine 2000 Transfection Reagent, Thermo Fisher Scientific, cat. no. 11668019*) according to the following protocol:

FuGENE / Lipofectamine 2000 protocol (PER WELL)

1) Tube A: 10µl Fugene + 160µl Optimem medium
2) Tube B:
   Test DNA 1.5µg
   psPAX2 1.3µg
   pMD2.G 0.5µg
   + 160µl Optimem
3) Incubate DNA + Optimem for 5 mins at room temperature

4) Using a P1000, gently add contents of tube A to tube B and mix
5) Incubate DNA/Fugene mixture for 30 mins at room temperature
6) Add all 320µl DNA/Fugene mixture to cells gently, & tilt plate to mix
7) Incubate cells for 48-72hrs before harvesting viral supernatant

Viral supernatants were harvested 48-72hrs after 293T cell transfection and filtered through a 0.2µm syringe filter (*Minisart Syringe Filter, Sartorius, cat. no. 16534-K*) to remove cell debris. Harvested virus was either used immediately for infection of target cells, or frozen in single-use aliquots at -80°C for later use.

Target cells were seeded for infection at a cell density of 200,000 cells per well in 6-well plates and allowed to adhere overnight. Viral supernatant was then added to the cells directly (100µl virus in 2ml of cell media per well of a 6-well plate), along with Polybrene Transfection Reagent (*Millipore, cat. no. TR-1003-G*) at a concentration of 8µg/ml to enhance the infection efficiency.

The amount of virus used for target cell infection was scaled up for larger cell culture formats as follows, based on a compromise between dish area and media volume (which would affect virus dilution):

Dish	Virus (µl)	Dish area (mm^2)	Media volume	Cells per well
6-well	100	962	2ml	200,000
10cm	350	7,854	7ml	$1x10^6$
15cm	1000	17,671	20ml	$2x10^6$

Subsequent antibiotic selection of infected cells was achieved with either puromycin, hygromycin, or blasticidin treatment (depending on the transduced construct in question). Puromycin (*InvivoGen, cat. no. ant-pr*) was diluted in cell culture media from a stock concentration of 10mg/ml to working concentrations of 3-4µg/ml. Hygromycin (*InvivoGen, cat. no. ant-hg*) was diluted in cell culture media from a stock concentration of 100mg/ml to working concentrations of 800-1000µg/ml. Blasticidin (*InvivoGen, cat. no. ant-bl*) was diluted in cell culture media from a stock concentration of 10mg/ml to working concentrations of 25µg/ml.

Virally infected target cells were treated with antibiotic alongside untransduced controls, and antibiotic treatment was stopped once all untransduced control cells were dead. This generally took 2-4 days in the case of puromycin, 5-7 days in the case of hygromycin, or 9 days on average in the case of blasticidin. During antibiotic selection, the antibiotic-supplemented medium was refreshed every 2-3 days and target cells were trypsinised and passaged or expanded as required. After antibiotic selection was complete, cells were cultured in fresh media (without antibiotics) for subsequent use.

FACs analysis and Cell sorting

FACs analysis was carried out using a BD LSRFortessa flow cytometer, and FACs data was analysed with FlowJo software v10. FACS sorting was carried out by the Flow Cytometry Core Facility at the Cambridge Institute for Medical Research.

Protein extraction & quantification

Adherent cells were typically cultured in 10cm dishes to obtain sufficient cells for protein extraction. Cells were first detached with 1X trypsin, washed off the plates with two media washes and collected in 15ml falcons. The cells were then pelleted by centrifugation (1200rpm, 5mins, 4°C) and either lysed to extract protein immediately or, in the case of time courses, frozen at -80°C for later extraction (so that all samples could be extracted at once, under the same conditions).

Cell pellets were then washed (i.e. resuspended) twice with ice cold PBS (and transferred to 1.5ml eppendorfs in the process), with a centrifugation step (1500rpm, 10mins, 4°C) to pellet the cells again after each wash. Pellets were then resuspended in ice cold RIPA lysis buffer (*Sigma, cat. no. R0278-50ML*), supplemented with Protease Inhibitor Cocktail (*Sigma, cat. no. P8340-1mL*) in a 1:100 dilution, and incubated on ice for 30mins. Protein lysates were then centrifuged at 14,000 rpm for 15mins at 4°C to remove any remaining cell debris, and the supernatant was transferred to a fresh, ice cold 1.5ml eppendorf for quantification and storage. Protein concentrations were measured via BCA assay with the Pierce BCA Protein Assay Kit (*Thermo Fisher Scientific, cat. no. 23227*) as per the kit instructions, and using the TECAN instrument to make absorbance measurements at a wavelength of 562nm. Protein lysate samples were then stored at -80°C for analysis by Western blot.

Genomic DNA extraction & quantification

Adherent cells were first detached with 1X trypsin, and pelleted by centrifugation at 1200rpm for 5mins at 4°C. Genomic DNA was then extracted from cell pellets with the QIAamp DNA Mini Kit (*Qiagen, cat. no. 51304*) as per the kit protocol. DNA concentrations were then measured using a Nanodrop. In the case of the screen samples, genomic DNA was extracted under bleached clean conditions in a separate room, and the DNA concentration was measured from a 2µl aliquot of the extracted sample, in order to prevent contamination from high copy DNA remaining from sgRNA plasmids or sgRNA cloning experiments. All genomic DNA samples were stored at -20°C.

RNA extraction & quantification

For qPCR, adherent cells were first detached with 1X trypsin, and pelleted by centrifugation (1200rpm, 5mins, 4°C). Total RNA was then extracted with RNAzol RT (*Sigma-Aldrich, cat. no. R4533*) according to the protocol recommended by the supplier.

For RNA-seq, adherent cells were detached with a cell scraper on ice, and pelleted by centrifugation (1200rpm, 5mins, 4°C). RNA-seq was performed with 786-M1A-Cas9 and OS-LM1B-Cas9 Control, sg2, and sg8 cells, with quadruplicate samples for each cell line. Thus, in order to minimise the influence of batch effects on the differences in mRNA levels between control and HNF1B knockout samples, the cells were scraped and pelleted in four batches, with each batch comprising a Control, sg2, and sg8 sample for either the 786-M1A-Cas9 cells or the OS-LM1B-Cas9 cells. Total RNA was then extracted using the Qiagen RNeasy Mini Kit (*Qiagen, cat. no. 74104*) according to the kit protocol. RNA concentrations were measured using a Nanodrop for all RNA samples (i.e. for both qPCR and RNA-seq samples). All RNA samples were then stored at -80°C.

Pelleting cells for RNA-seq:

1. Scrape the cells on ice as you would for protein pelleting/extraction
    a. Aspirate cell media, leaving 1-2ml of media behind
    b. Scrape the cells in media and transfer to 15ml falcon

c. Add another 1ml ice cold media and scrape a second time to recover more cells
d. Spin down (1000rpm, 5mins, 4°C)
e. Aspirate supernatant
f. Freeze pellet @ -80'C (snap-freeze on dry ice first, see below)

2. Be sure to scrape 1 plate at a time:
   a. Do one Control, one sg2, and one sg8 sample per batch
   b. Process one plate at a time: scrape and do steps 1a-1f, then move on to the next batch and repeat steps 1a-1f

      This will prevent a time delay between pelleting all four Control plates and all four sg2 or sg8 plates, which would minimise the risk of introducing differences in the RNA content of the cells between control and knockout samples groups. Instead, there will only be a time delay between the 1st batch of Control/sg2/sg8 and the 4th batch of Control/sg2/sg8 samples, so the experimental groups would be kept internally consistent within each batch.
   c. For step 1f, use a bucket with dry ice: transfer put pellets directly onto dry ice first, then transfer all falcons of pellets to -80°C.
3. Then extract RNA for all samples at once with RNeasy kit, as per kit protocol

Western blotting and antibodies

Western blots were performed using the Bio-Rad system with manually poured 10-well or 15-well SDS-PAGE gels (either 8% or 10% poly-acrylamide). In general, 10µg of protein sample was loaded per well, alongside a molecular weight marker (*Precision Plus Protein Kaleidoscope Standards, Bio-Rad, 1610375*). Gels were run at 80V, and transferred onto PVDF membranes (*Immobilon-P PVDF 0.45µm Membrane, Millipore, cat. no. IPVH00010*) for 2-3 hours at 110V. Membranes were blocked in a solution of 5% milk in 0.1% PBST. Primary and secondary antibody solutions were also made up in 5% milk in 0.1% PBST. The following primary antibodies were used:

Antibody	Supplier & Catalog Number	Antibody dilution and blotting conditions
mouse anti-β-actin	*Sigma, cat. no. A1978*	1:20,000, 4°C overnight

rabbit anti-HNF1β	*Human Protein Atlas, cat. no. HPA002083*	1:5000, 4°C overnight
mouse anti-Flag	*Sigma, cat. no. F1804*	1:500, 4°C overnight

The Flag antibody was used to blot for Cas9, which was linked to either a 1xFlag tag (in the constitutive lenti-Cas9-blast transgene) or a 3xFlag tag (in the conditional pCW-Cas9 transgene). The following HRP-conjugated secondary antibodies were used:

Antibody	**Supplier & Catalog Number**	**Antibody dilution and blotting conditions**
Polyclonal Goat Anti-Rabbit Immunoglobulins/HRP	*DAKO, cat. no. P0448*	1:5,000, 1hr at room temperature
Polyclonal Goat Anti-Mouse Immunoglobulins/HRP	*DAKO, cat. no. P0447*	1:10,000 30mins at room temperature for B-actin blots 1hr at room temperature for Flag blots
Rabbit Anti-Mouse IgG H&L (HRP)	*Abcam, cat. no. ab6728*	
Anti-rabbit IgG, HRP-linked Antibody	*Cell Signaling Technology, cat. no. 7074*	1:2,000, 1hr at room temperature

After each antibody incubation, membranes were washed with 0.1% PBST (1X PBS with 0.1% Tween-20, *Sigma, cat. no. T2700-100ML*) 3 times, for 5-10 mins per wash. For HNF1B, B-actin and Flag (specifically 3xFlag-Cas9), blots were developed with Luminata Classico Western HRP Substrate (*Millipore, cat. no. WBLUC0500*). For 1xFlag-Cas9, blots were developed with a high-sensitivity substrate, the Luminata Forte Western HRP Substrate (*Millipore, cat. no. WBLUF0100*). All blots were then imaged in a darkroom using X-Ray film (X Ray Film 18x24cm Double Sided; *Scientific Laboratory Supplies, cat. no. MOL7016*) and a film processor.

Quantitative reverse-transcriptase polymerase chain reaction (qRT-PCR)

Reverse transcriptase PCR (RT-PCR) was performed using the High-Capacity cDNA Reverse Transcription Kit (*Thermo Fisher Scientific, cat. no. 4368814*), with a modification of the recommended protocol:

2X RT Master Mix, w/out RNase inhibitor	µl
10X RT Buffer	2.00
25X dNTP Mix (100mM)	0.80
10X RT Random Primers	2.00
MultiScribe Reverse Transcriptase	0.25
Nuclease-free H_2O	4.95
Total volume per reaction:	**10.00**

500ng of RNA sample was used per reaction, in a total volume of 10µl. For each RT-PCR reaction, 10µl of RNA sample (a total of 500ng RNA) was added to the 10µl of 2X RT Master Mix, to give a total volume of 20µl (with all reagents at a 1X concentration). The following thermocycler programme was used (recommended by the kit protocol):

RT-PCR Programme:	
25°C	10 min
37°C	120 min
85°C	5 min
4°C	pause

The resultant cDNA (20µl total) was then diluted 1:10 with nuclease-free H_2O (by adding 180µl of nuclease-free H_2O), to give a final cDNA concentration of approximately 2.5ng/µl (assuming that all 500ng of RNA was reverse transcribed to give 500ng of cDNA). Quantitative (real-time) PCR (qPCR) was then performed by TaqMan Assay, using the TaqMan Fast Advanced Master Mix (*Thermo Fisher Scientific, cat. no. 4444557*) and the following TaqMan probes (*also from Thermo Fisher Scientific*):

Gene	Probe Name	Catalog Number
TBP (housekeeping gene)	TBP TaqMan probe, Hs00427620_m1 (size S: 250 rxns)	4331182
HNF1B	HNF1B TaqMan probe, Hs01001602_m1 (size S: 250 rxns)	4331182
MYC	MYC TaqMan probe, Hs00153408_m1, (size S: 250 rxns)	4331182
BCL2	BCL2 TaqMan probe, Hs00608023_m1, (size S: 250 rxns)	4331182
CDH16	CDH16 TaqMan probe, Hs00187880_m1, (size S: 250 rxns)	4331182

qPCR reactions were set up in 96-well qPCR plates (*96 Well Fast Plate, Sarstedt Ltd, cat. no. 72.1981.202*) as follows:

qPCR reaction:	μl
cDNA (1pg-100ng)	2.5
TaqMan Fast Advanced Master Mix (2X)	5.0
TaqMan probe (20X)	0.5
Nuclease-free H_2O	2.0
Total volume per reaction:	**10**

The qPCR plates were then sealed with optical adhesive film (*Adhesive Clear Foil for 96 Well Plate, Sarstedt Ltd, cat. no. 95.1994*) and run on a machine with the 7500 Fast Real-Time PCR System (*Applied Biosystems / Thermo Fisher Scientific*), using the following (machine default) qPCR programme:

1.	50°C	2 mins	
2.	95°C	20 secs	
3.	95°C	3 secs	x40 cycles
4.	60°C	30 secs	

The qPCR data was then analysed using the delta-delta-Ct (ΔΔCt) method to determine gene expression changes (specifically changes in mRNA levels in test samples relative to control samples).

RNA-seq

RNA-seq was performed with 786-M1A-Cas9 and OS-LM1B-Cas9 Control, sg2, and sg8 cells, with quadruplicate samples for each cell line. RNA was extracted from cell pellets using the Qiagen RNeasy Mini Kit (*Qiagen, cat. no. 74104*) according to the kit protocol, including the DNase I digestion step with the RNase-Free DNase Set (*Qiagen, cat. no. 79254*) to remove genomic DNA from the samples.

To avoid sample contamination with traces of RNA from other experiments, benches and equipment were first cleaned with RNase-Zap (*Ambion, Inc., cat. no. 9780*) before RNA extractions were carried out, fresh DNase/RNase-free distilled water (*Thermo Fisher Scientific, cat. no. 10977-035*) was used, and RNA samples were eluted into sterile, individually packaged polypropylene eppendorfs (*Microcentrifuge tube Biopur Safe-Lock individually sealed sterile 1.5 mL, Fisher Scientific, cat. no. 1051-8721*).

RNA concentrations were first measured using a Nanodrop. One aliquot was made for each sample for later use, containing 1000ng of RNA plus an extra 2µl of sample for Bioanalyzer analysis. This was in order to prevent multiple freeze-thaw cycles of the main sample stock. The aliquots and main sample stock were then stored at -80°C.

Quantification and quality checks for all of the RNA samples were then performed with a Bioanalyzer, using 2µl from the aliquot of each sample. Bioanalyzer analysis and quantification of the RNA samples was carried out using the Agilent RNA 6000 Nano Kit (*Agilent Technologies LDA UK Limited, cat. no. 5067-1511*), according to the kit instructions. Based on the Bioanalyzer concentrations for the RNA samples, a dilution was made for each sample to give a working concentration of 50ng/µl (i.e. an aliquot of each sample containing 500ng of RNA was diluted in RNase-free water up to a total volume of 10µl). These RNA sample dilutions were then used to prepare the RNA-seq libraries.

RNA-seq libraries were prepared by another member of the lab using the Lexogen SENSE mRNA-Seq Library Prep Kit V2 for Illumina, 96 preps (*Lexogen, cat. no. 001.96*),

according to the kit protocol. The final cDNA samples generated were then analysed and quantified with a Bioanalyzer using the Agilent High Sensitivity DNA Kit (*Agilent Technologies LDA UK Limited, cat. no. 5067-4626*). The RNA-seq libraries were then quantified by another member of the lab using the KAPA Library Quantification Kit (*KAPA Biosystems Ltd, cat. no. KK4835*). The RNA-seq libraries were then pooled and sent for Illumina sequencing.

The sequencing reads from the RNA-seq were aligned to hg38 using RSEM[128] and bowtie2[129] with default settings. The alignment of the RNA-seq sequencing reads to hg38, as well as the differential expression (DEseq2) analysis and gene set enrichment analysis (GSEA) of the RNA-seq sequencing data, were all performed in the lab in collaboration with bioinformaticians from the Samarjiwa lab. For both the 786-M1A-Cas9 and OS-LM1B-Cas9 sets of cell lines, the data for sg2 and sg8 was combined and then compared to the control using DEseq2, to identify changes that were significant and consistent between both sgRNAs. The gene expression changes were then compared between the 786-M1A-Cas9 and OS-LM1B-Cas9 datasets to discern which upregulated or downregulated genes were common to both cell lines.

ChIP-seq

HNF1B ChIP-seq was performed by another member of the lab, as previously published[133]. HNF1B ChIP was performed with the rabbit anti-HNF1B antibody from the Human Protein Atlas (*supplied by Sigma, cat. no. HPA002083*), with rabbit polyclonal IgG (*Abcam, ab27478*) used as a control. The ChIP protocol is briefly outlined below:

Chromatin immunoprecipitation (ChIP)

Obtaining cell pellet samples

1) Grow parental ccRCC cell lines to roughly 80% confluency in 15cm dishes.
2) Aspirate media from 15cm dishes and wash with PBS (7ml per dish).
3) Aspirate PBS and add trypsin (4ml per dish). Incubate cells at 37°C for 5mins or until cells detach.
4) Deactivate trypsin with media (10ml per dish). Transfer all cells +media +trypsin to 50ml falcon. Pool multiple 15cm dishes per cell line together in one falcon.
5) Spin down (1000rpm, 5mins, room temperature). Aspirate supernatant.

Pellet cross-linking

6) Cross-link cell pellets in 1% formaldehyde in RPMI media (10mins, room temperature)
7) Quench with 0.125 mol/L glycine (5mins, room temperature).
8) Spin down (1000rpm, 5mins, 4°C). Discard supernatant.
9) Wash pellets twice with ice cold 1X PBS. After each wash, remove supernatant and spin down (1000rpm, 5mins, 4°C).
10) Freeze cross-linked pellets at -80°C.

Immunoprecipitation (IP)

11) Vortex magnetic beads (*Pierce ChIP-grade Protein A/G Magnetic Beads, Thermo Fisher Scientific, cat. no. 26162*) to prevent precipitation.
12) Prepare one eppendorf per sample for each antibody.
13) Aliquot 100µl of beads per sample / IP into eppendorfs and put eppendorfs on ice.

IN COLD ROOM AT 4°C

14) Block magnetic beads by washing three times with 1ml 0.5% BSA in PBS.
15) Incubate magnetic beads with antibody in 0.5% BSA at 4°C for 4hrs minimum.
16) *AT ROOM TEMPERATURE*
17) Make up lysis buffer (20 mmol/L Tris–HCl pH 8.0, 150 mmol/L NaCl, 2 mmol/L EDTA pH 8.0, 0.1% SDS, and 1% Triton X-100)
18) Resuspend cross-linked cell pellets in lysis buffer and transfer cell suspension to a douncer.
19) Grind cells + lysis buffer in douncer to optimise cell lysis, and transfer lysate to eppendorfs on ice.
20) Incubate lysates on ice for 10mins.
21) Sonicate lysates with a Bioruptor machine (*Diagenode*) at high setting, 30s ON/OFF, for 14 cycles, in order to obtain fragments of 100-500bp in size.
22) Spin down sonicated samples (14000rpm, 20mins, 4°C). Retain supernatant and keep on ice.

IN COLD ROOM AT 4°C

23) Wash magnetic beads three times with 1ml 0.5% BSA, then resuspend beads in 100µl 0.5% BSA.
24) Add sonicated sample to antibody-coated beads and incubate at 4°C overnight (12-

16hrs).

25) Wash beads three times with low-salt buffer (50 mmol/L HEPES pH7.5, 140 mmol/L NaCl, 1% Triton)

26) Wash beads once with high-salt buffer (50 mmol/L HEPES pH7.5, 500 mmol/L NaCl, 1% Triton)

AT ROOM TEMPERATURE

27) Resuspend beads in elution buffer (50 mmol/L NaHCO3, 1% SDS).

28) Incubate at 65°C for 3hrs, shaking at 1000rpm, to de-crosslink samples.

29) Briefly spin down samples and place tubes in magnetic rack at room temperature. Allow solution to clear, and transfer supernatant to fresh eppendorf tube.

30) Purify DNA samples with the QuickClean II PCR Extraction Kit (*GenScript, cat. no. L00419-100*), according to the kit protocol. Elute twice with 50µl DNase/RNase-free distilled water, for a final eluate volume of 100µl.

31) Store DNA samples at -20°C.

32) Quantify samples with Qubit (as per Qubit quantification kit protocol).

The final DNA samples were analysed and quantified with a Bioanalyzer using the Agilent High Sensitivity DNA Kit (*Agilent Technologies LDA UK Limited, cat. no. 5067-4626*). ChIP-seq libraries were prepared by another member of the lab using the KAPA Hyper Prep Kit (*KAPA Biosystems Ltd, cat. no. KK8502*), according to the kit protocol. The ChIP-seq libraries were then quantified by another member of the lab using the KAPA Library Quantification Kit (*KAPA Biosystems Ltd, cat. no. KK4835*), and then pooled and sent for Illumina sequencing.

The HNF1B ChIP-seq sequencing data was analysed in the lab as follows: the raw ChIP-seq sequencing reads were aligned to hg38 using bowtie2[129], producing sam files, which were converted into sorted bam files using samtools[134]. HNF1B peaks were then called using MACS2[135] with the option -q 5e-2. The corresponding input samples from the ChIP experiment for each cell line were used as controls.

Any overlap between the peaks called for each cell line was then determined by taking the peak summits +/-50bp (bedops --range 50 --merge), and identifying regions that had a peak in at least two samples (i.e. in at least two out of the four cell lines); this was done using the multiIntersectBed function in bedtools. Thus, the final list of called peaks for HNF1B contained

peaks that were shared between at least two of the four samples. The gene annotation for the peaks was carried out using Homer[136]. The sequencing reads were first extended to the modal length of 250bp, and duplicates were removed; genome coverage was then calculated for each region of interest.

Transcription factor motif analysis was then carried out on the HNF1B peaks identified in the Illumina ChIP-seq sequencing data. First, the MEME suite module Meme[137] was used to carry out *de novo* motif discovery in the genomic sequences for the subset of HNF1B peaks that were shared by all four cell line samples. To perform this analysis, the summits +/- 50bp, from all the regions that had a peak in all four samples, were used to look for *de novo* motifs. The MEME suite module Tomtom[138] was then used to compare the motifs identified by the Meme analysis against known motifs.

CRISPR/Cas9 engineering (gene knockout)

sgRNA sequences were taken from the whole genome sgRNA library designed by Wang *et al*.[78]. The lentiviral sgRNA expression vector pKLV-U6-sgRNA(BbsI)-PGK-hygro-2A-BFP was cloned previously in the lab from the original pKLV-U6gRNA(BbsI)-PGKpuro2ABFP developed by Koike-Yusa *et al*.[79] (*Addgene plasmid ID: 50946*), by replacing the puromycin resistance gene with a hygromycin resistance gene. sgRNAs were cloned into pKLV-U6-sgRNA(BbsI)-PGK-hygro-2A-BFP as per the protocols outlined by Koike-Yusa *et al*.[79] and Ran *et al*.[127].

To generate stable conditional and constitutive knockout cell lines (used in the screen and for *in vitro* and *in vivo* assays), pKLV-U6-sgRNA(BbsI)-PGK-hygro-2A-BFP plasmids containing the desired sgRNAs were lentivirally transduced into ccRCC cell lines that expressed Cas9 under the control of either a dox-inducible or a constitutive promoter. Constitutive Cas9 cell lines were generated by transducing parental ccRCC cells with a constitutive Cas9 expression construct, lenti-Cas9-blast. Conditional (dox-inducible) Cas9 cell lines were generated by transducing parental ccRCC cells with a conditional Cas9 expression construct, pCW-Cas9.

Competition proliferation assays (with CRISPR/Cas9 cells)

For the original competition proliferation assays (*see Results section 2.2 and 2.3*), ccRCC cell lines (786-M1A, OS-LM1B, RFX631, RCC-MF, UOK101, and A498) and HeLa cells were transduced with the constitutive Cas9 expression construct (lenti-Cas9-blast), followed by one of the following sgRNAs: non-targeting control sgRNA #18 (NTC18, a.k.a Control), HNF1B sgRNA #2 (sg2), or HNF1B sgRNA #8 (sg8), in either an mCherry or BFP fluorescent backbone. The resultant cell lines (expressing Control-mCherry, Control-BFP, sg2-mCherry, sg2-BFP, sg8-mCherry, or sg8-BFP) were then used for the competition assays.

Cells expressing Control-mCherry were mixed in a 50:50 ratio with cells expressing either Control-BFP, sg2-BFP, or sg8-BFP. For the "colour swap" version of the competition assays, cells expressing Control-BFP were mixed in a 50:50 ratio with cells expressing either Control-mCherry, sg2-mCherry, or sg8-mCherry.

For the rescue competition assays (*see Results section 2.4*), 786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cell lines (already expressing the constitutive Cas9 expression construct lenti-Cas9-blast) were first transduced with either pLVX-puro(-) empty vector or pLVX-puro-HNF1B-rescue (i.e. the constitutive exogenous HNF1B rescue construct), and then transduced with either Control-BFP, Control-mCherry, sg8-BFP or sg8-mCherry. The resultant cell lines were then used for the rescue competition assays. For the first three experimental groups, Control-mCherry (+ empty vector) cells were mixed in a 50:50 ratio with either Control-BFP (+ empty vector), sg8-BFP (+ empty vector) or sg8-BFP (+ rescue) cells. In a fourth experimental group, sg8-mCherry (+ rescue) cells were mixed in a 50:50 ratio with sg8-BFP (+ empty vector) cells.

For the "colour swap" versions of the rescue competition assays, in the first three experimental groups, Control-BFP (+ empty vector) cells were mixed in a 50:50 ratio with either Control-mCherry (+ empty vector), sg8-mCherry (+ empty vector) or sg8-mCherry (+ rescue) cells. In the fourth experimental group, sg8-BFP (+ rescue) cells were mixed in a 50:50 ratio with sg8-mCherry (+ empty vector) cells.

For all of these CRISPR/Cas9 competition proliferation assays, the cell mixtures were cultured for 12-28 days. The percentages of BFP and mCherry positive cells were measured every 4 days by FACs, and the BFP:mCherry ratio (or the mCherry:BFP ratio in the case of the colour swap assays) was calculated for each time point (from Day 0 onwards). Thus, over the course of each assay, the cell mixtures were trypsinised every 4 days to take a sample for FACs

analysis, and passaged at a suitable ratio to be roughly 80% confluent in 4 days' time for the next time point.

During each CRISPR/Cas9 competition proliferation assay, the individual cell lines used in the assay were also cultured independently, to take samples for protein and/or RNA on Day 0 and at each subsequent 4-day assay time point. The individual cell lines were thus also trypsinised every 4 days and passaged at a suitable ratio to be roughly 80% confluent in 4 days' time for the next time point.

Competition proliferation assays (with CRISPRi cells)

All CRISPRi competition proliferation assays were performed in 786-M1A-dCas9-KRAB-mCherry and OS-LM1B-dCas9-KRAB-mCherry cells transduced with the relevant tandem sgRNA constructs (*cloned as described above*).

For the HNF1B CRISPRi (HNF1Bi) competition proliferation assays, untransduced parental BFP– cells (not transduced with sgRNAs) were mixed in a 50:50 ratio with BFP+ cells transduced with either the Scramble Control, HNF1B CRISPRi #1, HNF1B CRISPRi #2 or HNF1B CRISPRi #3 tandem sgRNA constructs.

For the MYC and BCL2 CRISPRi (MYCi/BCL2i) competition proliferation assays, untransduced parental BFP– cells (not transduced with sgRNAs) were mixed in a 50:50 ratio with transduced BFP+ cells, expressing either: Scramble Control, HNF1Bi (i.e. the HNF1B CRISPRi #2 tandem sgRNA construct), MYCi(1) alone, MYCi(2) alone, BCL2i alone, both MYCi(1) and BCL2i, or both MYCi(2) and BCL2i tandem sgRNA constructs.

For all of these CRISPRi competition proliferation assays, the cell mixtures were cultured for 16 days. The percentages of BFP positive versus BFP negative cells were measured every 4 days by FACs, and the BFP+/BFP– ratio was calculated for each time point (from Day 0 onwards). Thus, over the course of each assay, the cell mixtures were trypsinised every 4 days to take a sample for FACs analysis, and passaged at a suitable ratio to be roughly 80% confluent in 4 days' time for the next time point.

***In vitro* proliferation assays**

Cell counting and CellTiter-Glo proliferation assays were performed with 786-M1A-Cas9 and OS-LM1B-Cas9 Control-, sg2-, and sg8-BFP/mCherry cell lines.

For the cell counting assays, cells were seeded in triplicate in 6-well plates at a cell density of 10,000 cells per well. At each time point, cells were trypsinised and counted with Trypan Blue (*Life Technologies, cat. no. 15250061*) using Countess cell counting chamber slides (*Life Technologies, cat. no. C10228*) and the Countess Automated Cell Counter from Life Technologies.

For the CellTiter-Glo proliferation assays, cells were seeded in triplicate in 96-well plates at a cell density of 400 cells per well, alongside triplicate "blank" wells of media only (without cells) to provide values for background luminescence. The CellTiter-Glo proliferation assays were performed using the CellTiter-Glo Luminescent Cell Viability Assay kit (*Promega, cat. no. G7572*), with a modification of the protocol recommended by the supplier.

Both assays were carried out for 7 days; thus, for each cell line, separate sets of 6-well or 96-well wells were seeded for each day of measurements. The protocols for these two assays are shown below:

Day 0 + Day 1-7 measurements: *(N.B. D0 measurements taken 5hrs after cell seeding)*

CellTiter-Glo:

1. Equilibrate plates at room temperature for 30mins
2. Mix 50µl CellTiter-Glo reagent + 50µl DNase/RNase-free distilled water to 100µl of medium **per well**
3. Add 100µl reagent mix to 100µl medium in plate wells with a multichannel pipette
4. Pipette up and down to mix
5. Leave plates at room temp for 5mins (protected from light with foil)
6. Pipette up and down to mix again
7. Transfer entire 200µl from each well to an opaque-walled, clear-bottomed 96-well white plate (*Scientific Laboratory Supplies, cat. no. 3610*)
8. Incubate at room temp for 10mins (protected from light with foil)
9. Record luminescence signal with Tecan plate reader

Cell counting:

1. Trypsinise cells in 6-well plate (200µl trypsin)

2. Deactivate trypsin with 500µl medium per well
    a. pipette up and down to detach & mix cells
    b. transfer to falcon (use one falcon per well)
3. Add another 500µl medium to each well
    a. pipette up and down to wash off remaining cells
    b. add to same falcon for that well
4. Spin down (1200rpm, 5mins) and remove 1ml supernatant
5. Resuspend pellet in remaining 200µl
    a. take 50µl aliquot for Countess
    b. mix with 50µl trypan blue on parafilm
    c. pipette 12µl into Countess chamber slide
    d. count cells (cells/ml) with Countess Automated Cell Counter

***In vivo* tumour growth assays**

For subcutaneous tumour growth assays, control and HNF1B knockout (sg2 and sg8) 786-M1A-Cas9 and OS-LM1B-Cas9 ccRCC cell lines were resuspended in a 50/50 mixture of PBS/Matrigel (*Corning / Scientific Laboratory Supplies, cat. no. 356231*), at a concentration of $1x10^6$ cells/ml. Cells were injected subcutaneously into nude/athymic mice, into both the left and right flank of each mouse, with a total of $1x10^5$ cells in 100µl of cell suspension per injection site. 5 mice were used per cell line (i.e. per experimental group), with a total of 30 mice and 6 experimental groups (i.e. one group per cell line).

Weekly bioluminescence (luciferase-based) imaging of tumours was carried out using the IVIS Spectrum Xenogen machine (*Caliper Life Sciences*). D-luciferin solution was made up by dissolving 1g of D-luciferin powder (*D-Luciferin - K+ Salt; Perkin Elmer LAS (UK), cat. no. 122796*) in 66.667ml of sterile 1X PBS, to give a final concentration of 15mg/ml. For bioluminescence imaging, mice were injected intraperitoneally with 100µl D-luciferin (15mg/ml in 1X PBS), anaesthetized with isofluorane, and imaged with the Xenogen IVIS machine with different exposure times (10 secs, 20 secs, and auto-exposure). Bioluminescence measurements were quantified with the Living Image software (*Caliper Life Sciences*) to obtain the total photon flux per second (p/s) for the tumours in each mouse, at each weekly time point. The values for the total photon flux per second at each time point were then normalised to the corresponding

Day 0 values for each mouse. The normalised total photon flux per second served as an indicator of both the presence of tumour cells and a measure of tumour size *in vivo*. Changes in the normalised total photon flux per second over time were thus taken as a quantitative measure of tumour growth.

Weekly calliper measurements of tumour size were made once tumours became visible and palpable (which occurred in week 4 after injection). Calliper measurements were performed blind by the same individual every week to remove bias and minimise the effects of human error. Subcutaneous tumours were measured in two dimensions (width and length) in millimetres with callipers, and these measurements were used to calculate an approximation of tumour volume with the formula "$V = (L \times W^2) \times 0.5$" (where L is the length and W is the width of the tumour).

Experiments were terminated when the largest Control tumours reached a mean diameter of 12mm (which occurred after 7 weeks of tumour growth). Mice were then sacrificed by approved Schedule 1 methods (cervical dislocation), and subcutaneous tumours excised. All tumours were then weighed to determine the final mass of each tumour, and several tumours per experimental group were then fixed in 10% formalin for immunohistochemistry (IHC).

Immunohistochemistry

Excised subcutaneous tumours were fixed in 10% formalin for 24hrs. The formalin was then removed, and tumours were washed once with 1X PBS and transferred to 70% Ethanol. The fixed tumours were then used to make paraffin-embedded tissue blocks, which were cut into sections and placed on slides for immunohistochemistry (IHC; all courtesy of the Addenbrooke's Hospital Tissue Bank service). IHC stainings were carried out with the Leica Bond machine with the default protocol available in the Leica Bond software.

HNF1B IHC was performed with the rabbit anti-HNF1B antibody from the Human Protein Atlas (*supplied by Sigma, cat. no. HPA002083*) at an antibody concentration of 1:200. Vimentin IHC was performed with the rabbit anti-Vimentin antibody from Cell Signaling (*supplied by NEB, cat. no. 5741S*) at a 1:100 antibody dilution in Bond buffer
All antibodies were diluted in Bond buffer solution. Stained slides were then scanned and imaged at 10X zoom with the Zeiss Axioscan machine.

Propidium Iodide cell cycle assays

Cell cycle assays were performed with the 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines, and with the 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cell lines. Each cell line was seeded into eleven 6cm dishes: one dish per assay time point (every 2 hours from t0-t16, and a tenth time point at t24), plus one dish for unsynchronised cells (i.e. untreated, U/T). Cells were seeded at a suitable cell density such that they would be 70-80% confluent the following day, and allowed to adhere overnight.

The following morning, cells in the t0-t24 time point dishes were treated with 50ng/ml nocodazole for 24 hours, while the cells in the U/T dishes were left to grow untreated as a control. Nocodazole treatment served to arrest the cells in the G2/M phase of the cell cycle, thus synchronising the cell cycles of the different cell lines.

At the end of the 24 hour treatment period, treated cells were taken off nocodazole to release the synchronised cells from mitotic arrest. This was done by media changing to fresh media without nocodazole. Cells were released from nocodazole (i.e. grown in fresh media) for another 24 hours. During this period, samples (cell pellets) were taken for the U/T unsynchronised cells and for each cell line immediately after release (i.e. the t0 time point), and then for each cell line every 2 hours for 16 hours (the t2-t16 time points) and at 24 hours post-release (the t24 time point).

To take samples at each of these time points, cells were pelleted entirely from the dish corresponding to the appropriate time point. Immediately after pelleting the cells, each cell pellet sample was fixed in 500µl 70% EtOH and frozen at -20°C. The samples were then stained with Propidium Iodide (PI), and the cell cycle distribution at each time point was analysed by FACs. The protocol for pelleting the cells and fixing and staining the samples with PI was as follows:

Propidium Iodide Cell Cycle Staining

Obtaining cell pellet samples

1) Pipette cell media up and down in dish to wash off cells in mitotic arrest. Transfer media to 15ml falcon.
2) Add PBS to wash the cells. Pipette up and down and transfer PBS to same falcon.
3) Add trypsin to detach remaining cells. Incubate at 37°C until cells detached.
4) Deactivate trypsin with media and transfer all cells +media +trypsin to same falcon.

5) Wash dish with media to remove any remaining trypsinised cells. Transfer cells +media to same falcon.
6) Spin down (1000rpm, 5mins, room temperature).
7) Remove supernatant.

Sample fixation *(N.B. fix samples one by one)*

8) Resuspend cell pellets in 150µl 1X PBS
9) Add 350µl ice-cold 100% EtOH, vortex to mix, and put falcons with cell pellets on ice.
10) Store falcons (i.e. fixed cell pellets) at -20°C overnight (or until required for PI staining).

Propidium Iodide staining

11) Make up blocking solution: 1% BSA in PBS (1g BSA powder in 100ml 1X PBS).
12) Make up PI in blocking solution to a final concentration of 40µg/ml Propidium Iodide (*Sigma, cat. no. P4864-10ML*).
13) Add RNase A to PI + blocking solution, to a final concentration of 200µg/ml RNase A (*Sigma, cat. no. 10109169001*).
14) Keep PI/RNase A solution (i.e. 40µg/ml PI + 200µg/ml RNase A in blocking solution) wrapped in foil, protected from light.
15) Add 5ml blocking solution (1% BSA in PBS) to each sample / falcon of fixed cells.
16) Spin down (1000rpm, 5mins, room temperature).
17) Aspirate supernatant.
18) Resuspend cell pellets in 750µl PI/RNase A solution.
19) Transfer cells to polystyrene FACs tubes (*Scientific Laboratory Supplies, cat. no. 352054*) wrapped in foil.
20) Incubate samples in the dark for 30mins at room temperature.
21) Analyse samples by FACs.

Annexin V / PI apoptosis assays

Annexin V / PI assays were performed with the 786-M1A-Cas9 Control-BFP, sg2-BFP, and sg8-BFP cell lines, and with the 786-M1A-Cas9 Control-BFP (+ empty vector), sg8-BFP (+ empty vector) and sg8-BFP (+ rescue) cell lines. Annexin V / PI staining was carried out using the

eBioscience Annexin V Apoptosis Detection Kit APC (*Thermo Fisher Scientific, cat. no. 88-8007*), with APC-conjugated Annexin V.

Cells were seeded into 6cm dishes in normal media (RPMI +10% FCS +1% P/S) and allowed to adhere. Annexin V / PI staining was then carried out with a modification of the kit protocol, as follows:

Annexin V / PI apoptosis staining

1) Pre-cool a 15ml falcon on ice.
2) Pipette off cell media and transfer to falcon.
3) Add 2ml PBS to wash the cells. Pipette off PBS and transfer to same falcon.
4) Add 1ml trypsin to detach cells. Incubate at 37°C until cells detached.
5) Deactivate trypsin with media and transfer all cells +media +trypsin to same falcon.
6) Wash dish with media to remove any remaining trypsinised cells. Transfer cells +media to same falcon.
7) Spin down (2000rpm, 5mins, 4°C).
8) Remove supernatant. Gently resuspend cell pellet in 1ml cold PBS.
9) Add another 4ml cold PBS to this cell suspension. Spin down again (2000rpm, 5mins, 4°C). Repeat this wash step 2-3 times to remove all trypsin.
10) Dilute 10X Binding Buffer to 1X using distilled water (1ml 10X Binding Buffer + 9ml dH_2O).
11) Remove supernatant and gently resuspend cell pellet in **100µl 1X Binding Buffer** (i.e. for final cell concentration between $1x10^6$ to $5x10^6$ cells/ml). Transfer cells to fresh 1.5ml eppendorf.
12) Add **5µl of Annexin V stain** to each sample (i.e. add 5µl of fluorochrome-conjugated Annexin V to 100µl of the cell suspension).
13) Then add **5µl of propidium iodide stain** to each sample.
14) Incubate samples on ice in the dark for 15mins.
15) After incubation with both stains, add **320µl 1X Binding Buffer** to each sample.
16) Analyse on FACs immediately (within 4hrs; if storing samples, store @ 4'C in the dark).

REFERENCES

1. Motzer, R. J., Bander, N. H. & Nanus, D. M. Renal-Cell Carcinoma. *N. Engl. J. Med.* **335,** 865–875 (1996).
2. Cohen, H. T. & McGovern, F. J. Renal-Cell Carcinoma. *N. Engl. J. Med.* **353,** 2477–2490 (2005).
3. Chuang, S. *et al.* Overexpression of Glutathione S-Transferase α in Clear Cell Renal Cell Carcinoma. *Am. J. Clin. Pathol.* **123,** 421–429 (2005).
4. Vanharanta, S. *et al.* Epigenetic expansion of VHL-HIF signal output drives multiorgan metastasis in renal cancer. *Nat. Med.* **19,** 50–56 (2013).
5. Schlesinger-Raab, A., Treiber, U., Zaak, D., Hölzel, D. & Engel, J. Metastatic renal cell carcinoma: Results of a population-based study with 25 years follow-up. *Eur. J. Cancer* **44,** 2485–2495 (2008).
6. Hess, K. R. *et al.* Metastatic patterns in adenocarcinoma. *Cancer* **106,** 1624–1633 (2006).
7. Renshaw, A. A. & Richie, J. P. Subtypes of renal cell carcinoma. Different onset and sites of metastatic disease. *Am. J. Clin. Pathol.* **111,** 539–43 (1999).
8. Huang, D. *et al.* Inhibition of MAPK Kinase Signaling Pathways Suppressed Renal Cell Carcinoma Growth and Angiogenesis In vivo. *Cancer Res.* **68,** 81–88 (2008).
9. Motzer, R. J. & Russo, P. Systemic therapy for renal cell carcinoma. *J. Urol.* **163,** 408–17 (2000).
10. Rini, B. I., Campbell, S. C., Escudier, B. & Roussy, I. G. Renal cell carcinoma. *Lancet* **373,** 1119–1132 (2009).
11. Grandinetti, C. A. & Goldspiel, B. R. Sorafenib and Sunitinib: Novel Targeted Therapies for Renal Cell Cancer. *Pharmacotherapy* **27,** 1125–1144 (2007).
12. Rathmell, W. K., Martz, C. A. & Rini, B. I. Renal cell carcinoma. *Curr. Opin. Oncol.* **19,** 234–240 (2007).
13. Escudier, B., Goupil, M. G., Massard, C. & Fizazi, K. Sequential therapy in renal cell carcinoma. *Cancer* **115,** 2321–2326 (2009).
14. Motzer, R. J. Renal Cell Carcinoma: A Priority Malignancy for Development and Study of Novel Therapies. *J. Clin. Oncol.* **21,** 1193–1194 (2003).
15. Nanus, D. M., Pfeffer, L. M., Bander, N. H., Bahri, S. & Albino, A. P. Antiproliferative

and antitumor effects of alpha-interferon in renal cell carcinomas: correlation with the expression of a kidney-associated differentiation glycoprotein. *Cancer Res.* **50,** 4190–4 (1990).

16. Fyfe, G. *et al.* Results of treatment of 255 patients with metastatic renal cell carcinoma who received high-dose recombinant interleukin-2 therapy. *J. Clin. Oncol.* **13,** 688–696 (1995).
17. Fisher, R. I., Rosenberg, S. A. & Fyfe, G. Long-term survival update for high-dose recombinant interleukin-2 in patients with renal cell carcinoma. *Cancer J. Sci. Am.* **6 Suppl 1,** S55-7 (2000).
18. Yang, J. C. *et al.* Randomized Study of High-Dose and Low-Dose Interleukin-2 in Patients With Metastatic Renal Cancer. *J. Clin. Oncol.* **21,** 3127–3132 (2003).
19. McDermott, D. F. *et al.* Randomized Phase III Trial of High-Dose Interleukin-2 Versus Subcutaneous Interleukin-2 and Interferon in Patients With Metastatic Renal Cell Carcinoma. *J. Clin. Oncol.* **23,** 133–141 (2005).
20. Rosenberg, S. A., Yang, J. C., White, D. E. & Steinberg, S. M. Durability of complete responses in patients with metastatic cancer treated with high-dose interleukin-2: identification of the antigens mediating response. *Ann. Surg.* **228,** 307–19 (1998).
21. McDermott, D. F. *et al.* Survival, Durable Response, and Long-Term Safety in Patients With Previously Treated Advanced Renal Cell Carcinoma Receiving Nivolumab. *J. Clin. Oncol.* **33,** 2013–2020 (2015).
22. Motzer, R. J. *et al.* Nivolumab for Metastatic Renal Cell Carcinoma: Results of a Randomized Phase II Trial. *J. Clin. Oncol.* **33,** 1430–1437 (2015).
23. Motzer, R. J. *et al.* Nivolumab versus Everolimus in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **373,** 1803–1813 (2015).
24. Kaelin, W. G. Molecular basis of the VHL hereditary cancer syndrome. *Nat. Rev. Cancer* **2,** 673–682 (2002).
25. Kim, W. Y. & Kaelin, W. G. Role of VHL Gene Mutation in Human Cancer. *J. Clin. Oncol.* **22,** 4991–5004 (2004).
26. Fabbro, D., Parkinson, D. R. & Matter, A. Protein tyrosine kinase inhibitors: new treatment modalities? *Curr. Opin. Pharmacol.* **2,** 374–381 (2002).
27. Keshet, E. & Ben-Sasson, S. A. Anticancer drug targets: approaching angiogenesis. *J.*

Clin. Invest. **104,** 1497–1501 (1999).

28. Drevs, J., Medinger, M., Schmidt-Gersbach, C., Weber, R. & Unger, C. Receptor tyrosine kinases: the main targets for new anticancer therapy. *Curr. Drug Targets* **4,** 113–21 (2003).
29. Ratain, M. J. *et al.* Phase II Placebo-Controlled Randomized Discontinuation Trial of Sorafenib in Patients With Metastatic Renal Cell Carcinoma. *J. Clin. Oncol.* **24,** 2505–2512 (2006).
30. Kane, R. C. *et al.* Sorafenib for the Treatment of Advanced Renal Cell Carcinoma. *Clin. Cancer Res.* **12,** 7271–7278 (2006).
31. Escudier, B. *et al.* Sorafenib in Advanced Clear-Cell Renal-Cell Carcinoma. *N. Engl. J. Med.* **356,** 125–134 (2007).
32. Rini, B. I. *et al.* Comparative effectiveness of axitinib versus sorafenib in advanced renal cell carcinoma (AXIS): a randomised phase 3 trial. *Lancet* **378,** 1931–1939 (2011).
33. Goodman, V. L. *et al.* Approval Summary: Sunitinib for the Treatment of Imatinib Refractory or Intolerant Gastrointestinal Stromal Tumors and Advanced Renal Cell Carcinoma. *Clin. Cancer Res.* **13,** 1367–1373 (2007).
34. Rock, E. P. *et al.* Food and Drug Administration Drug Approval Summary: Sunitinib Malate for the Treatment of Gastrointestinal Stromal Tumor and Advanced Renal Cell Carcinoma. *Oncologist* **12,** 107–113 (2007).
35. Motzer, R. J. *et al.* Sunitinib versus Interferon Alfa in Metastatic Renal-Cell Carcinoma. *N. Engl. J. Med.* **356,** 115–124 (2007).
36. Motzer, R. J. *et al.* Overall Survival and Updated Results for Sunitinib Compared With Interferon Alfa in Patients With Metastatic Renal Cell Carcinoma. *J. Clin. Oncol.* **27,** 3584–3590 (2009).
37. Motzer, R. J. *et al.* Sunitinib Efficacy Against Advanced Renal Cell Carcinoma. *J. Urol.* **178,** 1883–1887 (2007).
38. Yang, J. C. *et al.* A Randomized Trial of Bevacizumab, an Anti–Vascular Endothelial Growth Factor Antibody, for Metastatic Renal Cancer. *N. Engl. J. Med.* **349,** 427–434 (2003).
39. Escudier, B. *et al.* Bevacizumab plus interferon alfa-2a for treatment of metastatic renal cell carcinoma: a randomised, double-blind phase III trial. *Lancet* **370,** 2103–2111 (2007).

40. Drevs, J. *et al.* Effects of PTK787/ZK 222584, a specific inhibitor of vascular endothelial growth factor receptor tyrosine kinases, on primary tumor, metastasis, vessel density, and blood flow in a murine renal cell carcinoma model. *Cancer Res.* **60,** 4819–24 (2000).
41. Rowinsky, E. K. *et al.* Safety, Pharmacokinetics, and Activity of ABX-EGF, a Fully Human Anti–Epidermal Growth Factor Receptor Monoclonal Antibody in Patients With Metastatic Renal Cell Cancer. *J. Clin. Oncol.* **22,** 3003–3015 (2004).
42. Prewett, M. *et al.* Mouse-human chimeric anti-epidermal growth factor receptor antibody C225 inhibits the growth of human renal cell carcinoma xenografts in nude mice. *Clin. Cancer Res.* **4,** 2957–66 (1998).
43. Atkins, M. B. *et al.* Randomized Phase II Study of Multiple Dose Levels of CCI-779, a Novel Mammalian Target of Rapamycin Kinase Inhibitor, in Patients With Advanced Refractory Renal Cell Carcinoma. *J. Clin. Oncol.* **22,** 909–918 (2004).
44. Hudes, G. *et al.* Temsirolimus, Interferon Alfa, or Both for Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* **356,** 2271–2281 (2007).
45. Motzer, R. J. *et al.* Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet* **372,** 449–456 (2008).
46. Bhatia, S. & Thompson, J. A. Temsirolimus in patients with advanced renal cell carcinoma: an overview. *Adv. Ther.* **26,** 55–67 (2009).
47. Rini, B. I. & Atkins, M. B. Resistance to targeted therapy in renal-cell carcinoma. *Lancet Oncol.* **10,** 992–1000 (2009).
48. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nat. Rev. Urol.* **7,** 277–285 (2010).
49. Network, T. C. G. A. R. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499,** 43–49 (2013).
50. Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell renal cell carcinomas defined by multiregion sequencing. *Nat. Genet.* **46,** 225–233 (2014).
51. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
52. Guo, G. *et al.* Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat. Genet.* **44,** 17–9 (2011).
53. Gao, J. *et al.* Integrative Analysis of Complex Cancer Genomics and Clinical Profiles

Using the cBioPortal. *Sci. Signal.* **6,** pl1-pl1 (2013).

54. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov.* **2,** 401–404 (2012).
55. Kondo, K., Klco, J., Nakamura, E., Lechpammer, M. & Kaelin, W. G. Inhibition of HIF is necessary for tumor suppression by the von Hippel-Lindau protein. *Cancer Cell* **1,** 237–46 (2002).
56. Ohh, M. *et al.* Synthetic peptides define critical contacts between elongin C, elongin B, and the von Hippel-Lindau protein. *J. Clin. Invest.* **104,** 1583–1591 (1999).
57. Ohh, M. *et al.* Ubiquitination of hypoxia-inducible factor requires direct binding to the β-domain of the von Hippel–Lindau protein. *Nat. Cell Biol.* **2,** 423–427 (2000).
58. Stebbins, C. E., Kaelin, W. G. & Pavletich, N. P. Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284,** 455–61 (1999).
59. Gossage, L. *et al.* An integrated computational approach can classify VHL missense mutations according to risk of clear cell renal carcinoma. *Hum. Mol. Genet.* **23,** 5976–5988 (2014).
60. Gossage, L. & Eisen, T. Alterations in VHL as potential biomarkers in renal-cell carcinoma. *Nat. Rev. Clin. Oncol.* **7,** 277–288 (2010).
61. Gossage, L., Eisen, T. & Maher, E. R. VHL, the story of a tumour suppressor gene. *Nat. Rev. Cancer* **15,** 55–64 (2015).
62. Forman, J. R., Worth, C. L., Bickerton, G. R. J., Eisen, T. G. & Blundell, T. L. Structural bioinformatics mutation analysis reveals genotype-phenotype correlations in von Hippel-Lindau disease and suggests molecular mechanisms of tumorigenesis. *Proteins Struct. Funct. Bioinforma.* **77,** 84–96 (2009).
63. Mandriota, S. J. *et al.* HIF activation identifies early lesions in VHL kidneys: evidence for site-specific tumor suppressor function in the nephron. *Cancer Cell* **1,** 459–68 (2002).
64. Kondo, K., Kim, W. Y., Lechpammer, M. & Kaelin, W. G. Inhibition of HIF2α is sufficient to suppress pVHL-defective tumor growth. *PLoS Biol.* **1,** 439–444 (2003).
65. Iliopoulos, O., Kibel, A., Gray, S. & Kaelin, W. G. Tumour suppression by the human von Hippel-Lindau gene product. *Nat. Med.* **1,** 822–826 (1995).
66. Baba, M. *et al.* Tumor suppressor protein VHL is induced at high cell density and

mediates contact inhibition of cell growth. *Oncogene* **20,** 2727–2736 (2001).

67. Chen, F. *et al.* Suppression of growth of renal carcinoma cells by the von Hippel-Lindau tumor suppressor gene. *Cancer Res.* **55,** 4804–7 (1995).
68. Davidowitz, E. J., Schoenfeld, A. R. & Burk, R. D. VHL Induces Renal Cell Differentiation and Growth Arrest through Integration of Cell-Cell and Cell-Extracellular Matrix Signaling. *Mol. Cell. Biol.* **21,** 865–874 (2001).
69. Lieubeau-Teillet, B. *et al.* von Hippel-Lindau gene-mediated growth suppression and induction of differentiation in renal cell carcinoma cells grown as multicellular tumor spheroids. *Cancer Res.* **58,** 4957–62 (1998).
70. Pause, A., Lee, S., Lonergan, K. M. & Klausner, R. D. The von Hippel-Lindau tumor suppressor gene is required for cell cycle exit upon serum withdrawal. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 993–8 (1998).
71. Zimmer, M., Doucette, D., Siddiqui, N. & Iliopoulos, O. Inhibition of hypoxia-inducible factor is sufficient for growth suppression of VHL-/- tumors. *Mol. Cancer Res.* **2,** 89–95 (2004).
72. Maranchie, J. K. *et al.* The contribution of VHL substrate binding and HIF1-alpha to the phenotype of VHL loss in renal cell carcinoma. *Cancer Cell* **1,** 247–55 (2002).
73. Gnarra, J. R. *et al.* Defective placental vasculogenesis causes embryonic lethality in VHL-deficient mice. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 9102–7 (1997).
74. Haase, V. H., Glickman, J. N., Socolovsky, M. & Jaenisch, R. Vascular tumors in livers with targeted inactivation of the von Hippel-Lindau tumor suppressor. *Proc. Natl. Acad. Sci.* **98,** 1583–1588 (2001).
75. Cho, H. *et al.* On-target efficacy of a HIF-2α antagonist in preclinical kidney cancer models. *Nature* **539,** 107–111 (2016).
76. Chen, W. *et al.* Targeting renal cell carcinoma with a HIF-2 antagonist. *Nature* **539,** 112–117 (2016).
77. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science (80-.).* **343,** 84–87 (2014).
78. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science (80-.).* **343,** 80–84 (2014).
79. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide

recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32,** 267–273 (2014).

80. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation. *Annu. Rev. Genet.* **45,** 273–297 (2011).
81. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).* **339,** 819–823 (2013).
82. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).
83. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science (80-.).* **337,** 816–821 (2012).
84. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153,** 910–918 (2013).
85. Yang, H. *et al.* One-Step Generation of Mice Carrying Reporter and Conditional Alleles by CRISPR/Cas-Mediated Genome Engineering. *Cell* **154,** 1370–1379 (2013).
86. Darnell, J. E. Transcription factors as targets for cancer therapy. *Nat. Rev. Cancer* **2,** 740–749 (2002).
87. Bommi-Reddy, A. *et al.* Kinase requirements in human cells: III. Altered kinase requirements in VHL-/- cancer cells detected in a pilot synthetic lethal screen. *Proc. Natl. Acad. Sci.* **105,** 16484–16489 (2008).
88. Ding, Y. *et al.* Combined Gene Expression Profiling and RNAi Screening in Clear Cell Renal Cell Carcinoma Identify PLK1 and Other Therapeutic Kinase Targets. *Cancer Res.* **71,** 5225–5234 (2011).
89. Chan, D. A. *et al.* Targeting GLUT1 and the Warburg Effect in Renal Cell Carcinoma by Chemical Synthetic Lethality. *Sci. Transl. Med.* **3,** 94ra70-94ra70 (2011).
90. Gerlinger, M. *et al.* Genome-wide RNA interference analysis of renal carcinoma survival regulators identifies MCT4 as a Warburg effect metabolic target. *J. Pathol.* **227,** 146–156 (2012).
91. Takahashi, K. & Yamanaka, S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell* **126,** 663–676 (2006).
92. Gronych, J., Pfister, S. M. & Jones, D. T. W. Connect Four with Glioblastoma Stem Cell

Factors. *Cell* **157,** 525–527 (2014).

93. Suvà, M. L. *et al.* Reconstructing and Reprogramming the Tumor-Propagating Potential of Glioblastoma Stem-like Cells. *Cell* **157,** 580–594 (2014).
94. Rheinbay, E. *et al.* An Aberrant Transcription Factor Network Essential for Wnt Signaling and Stem Cell Maintenance in Glioblastoma. *Cell Rep.* **3,** 1567–1579 (2013).
95. Johnston, S. J. & Carroll, J. S. Transcription factors and chromatin proteins as therapeutic targets in cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1855,** 183–192 (2015).
96. Johannessen, C. M. *et al.* A melanocyte lineage program confers resistance to MAP kinase pathway inhibition. *Nature* **504,** 138–142 (2013).
97. Clissold, R. L., Hamilton, A. J., Hattersley, A. T., Ellard, S. & Bingham, C. HNF1B-associated renal and extra-renal disease—an expanding clinical spectrum. *Nat. Rev. Nephrol.* **11,** 102–112 (2015).
98. Bockenhauer, D. & Jaureguiberry, G. HNF1B-associated clinical phenotypes: the kidney and beyond. *Pediatr. Nephrol.* **31,** 707–714 (2015).
99. Verhave, J. C., Bech, A. P., Wetzels, J. F. M. & Nijenhuis, T. Hepatocyte Nuclear Factor 1β-Associated Kidney Disease: More than Renal Cysts and Diabetes. *J. Am. Soc. Nephrol.* **27,** 345–353 (2016).
100. Rebouissou, S. *et al.* Germline hepatocyte nuclear factor 1α and 1β mutations in renal cell carcinomas. *Hum. Mol. Genet.* **14,** 603–614 (2005).
101. Hiesberger, T. *et al.* Role of the Hepatocyte Nuclear Factor-1β (HNF-1β) C-terminal Domain in Pkhd1 (ARPKD) Gene Transcription and Renal Cystogenesis. *J. Biol. Chem.* **280,** 10578–10586 (2005).
102. Suzuki, E. *et al.* Transcriptional upregulation of HNF-1β by NF-κB in ovarian clear cell carcinoma modulates susceptibility to apoptosis through alteration in bcl-2 expression. *Lab. Investig.* **95,** 962–972 (2015).
103. Hajarnis, S. S. *et al.* Transcription Factor Hepatocyte Nuclear Factor-1β (HNF-1β) Regulates MicroRNA-200 Expression through a Long Noncoding RNA. *J. Biol. Chem.* **290,** 24793–24805 (2015).
104. Wang, C., Mao, T., Yang, W. & Jeng, Y. Underexpression of hepatocyte nuclear factor-1β in chromophobe renal cell carcinoma. *Histopathology* **62,** 589–594 (2013).
105. Coffinier, C., Thépot, D., Babinet, C., Yaniv, M. & Barra, J. Essential role for the

homeoprotein vHNF1/HNF1beta in visceral endoderm differentiation. *Development* **126,** 4785–94 (1999).

106. Barbacci, E. *et al.* Variant hepatocyte nuclear factor 1 is required for visceral endoderm specification. *Development* **126,** 4795–805 (1999).
107. Hiesberger, T. *et al.* Mutation of hepatocyte nuclear factor-1beta inhibits Pkhd1 gene expression and produces renal cysts in mice. *J. Clin. Invest.* **113,** 814–25 (2004).
108. Gresh, L. *et al.* A transcriptional network in polycystic kidney disease. *EMBO J.* **23,** 1657–1668 (2004).
109. Kaminski, M. M. *et al.* Direct reprogramming of fibroblasts into renal tubular epithelial cells by defined transcription factors. *Nat. Cell Biol.* **18,** 1269–1280 (2016).
110. Mandai, M. *et al.* Ovarian clear cell carcinoma meets metabolism; HNF-1β; confers survival benefits through the Warburg effect and ROS reduction. *Oncotarget* **6,** (2015).
111. Gounaris, I. & Brenton, J. D. Molecular pathogenesis of ovarian clear cell carcinoma. *Futur. Oncol.* **11,** 1389–1405 (2015).
112. Okamoto, T. *et al.* Hepatocyte nuclear factor-1β (HNF-1β) promotes glucose uptake and glycolytic activity in ovarian clear cell carcinoma. *Mol. Carcinog.* **54,** 35–49 (2015).
113. Zhao, Y. *et al.* Meta-analysis of the association between the HNF1B rs4430796 (A>G) polymorphism and risk of prostate cancer based on case-control studies. *Genet. Mol. Res.* **14,** 7426–7435 (2015).
114. Xiang, Y. Z. *et al.* Racial disparities in the association between diabetes mellitus-associated polymorphic locus rs4430796 of the HNF1βgene and prostate cancer: a systematic review and meta-analysis. *Genet. Mol. Res.* **13,** 6582–6592 (2014).
115. Debiais-Delpech, C. *et al.* Expression patterns of candidate susceptibility genes HNF1β and CtBP2 in prostate cancer: Association with tumor progression. *Urol. Oncol. Semin. Orig. Investig.* **32,** 426–432 (2014).
116. Painter, J. N. *et al.* Fine-mapping of the HNF1B multicancer locus identifies candidate variants that mediate endometrial cancer risk. *Hum. Mol. Genet.* **24,** 1478–1492 (2015).
117. Tsuchiya, A. *et al.* Expression profiling in ovarian clear cell carcinoma: identification of hepatocyte nuclear factor-1 beta as a molecular marker and a possible molecular target for therapy of ovarian clear cell carcinoma. *Am. J. Pathol.* **163,** 2503–12 (2003).
118. Spurdle, A. B. *et al.* Genome-wide association study identifies a common variant

associated with risk of endometrial cancer. *Nat. Genet.* **43,** 451–454 (2011).

119. Gudmundsson, J. *et al.* Two variants on chromosome 17 confer prostate cancer risk and the one in TCF2 protects against type 2 diabetes. *Nat. Genet.* **39,** 977–983 (2007).
120. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40,** 310–315 (2008).
121. Sun, J. *et al.* Evidence for two independent prostate cancer risk–associated loci in the HNF1B gene at 17q12. *Nat. Genet.* **40,** 1153–1155 (2008).
122. Ponomarev, V. *et al.* A novel triple-modality reporter gene for whole-body fluorescent, bioluminescent, and nuclear noninvasive imaging. *Eur. J. Nucl. Med. Mol. Imaging* **31,** 740–751 (2004).
123. Zuber, J. *et al.* Toolkit for evaluating genes required for proliferation and survival using tetracycline-regulated RNAi. *Nat. Biotechnol.* **29,** 79–83 (2011).
124. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
125. Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154,** 442–51 (2013).
126. Gilbert, L. A. *et al.* Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159,** 647–61 (2014).
127. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8,** 2281–2308 (2013).
128. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
129. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–9 (2012).
130. Gabay, M., Li, Y. & Felsher, D. W. MYC Activation Is a Hallmark of Cancer Initiation and Maintenance. *Cold Spring Harb. Perspect. Med.* **4,** a014241–a014241 (2014).
131. Dang, C. V. MYC on the Path to Cancer. *Cell* **149,** 22–35 (2012).
132. Bretones, G., Delgado, M. D. & León, J. Myc and cell cycle control. *Biochim. Biophys. Acta - Gene Regul. Mech.* **1849,** 506–516 (2015).
133. Rodrigues, P. *et al.* NF-κB-Dependent Lymphoid Enhancer Co-option Promotes Renal Carcinoma Metastasis. *Cancer Discov.* **8,** 850–865 (2018).

134. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
135. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
136. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38,** 576–89 (2010).
137. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* **2,** 28–36 (1994).
138. Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. & Noble, W. S. Quantifying similarity between motifs. *Genome Biol.* **8,** R24 (2007).
139. Vanharanta, S. & Massagué, J. Origins of Metastatic Traits. *Cancer Cell* **24,** 410–421 (2013).
140. Cho, S. W. *et al.* Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. *Cell* **173,** 1398–1412.e22 (2018).
141. Cui, M. *et al.* Long non-coding RNA PVT1 and cancer. *Biochem. Biophys. Res. Commun.* **471,** 10–14 (2016).
142. Houshmand, M. *et al.* Long non-coding RNA PVT1 as a novel candidate for targeted therapy in hematologic malignancies. *Int. J. Biochem. Cell Biol.* **98,** 54–64 (2018).
143. Marchese, F. P. & Huarte, M. A 'Counter-Enhancer' in Tumor Suppression. *Cell* **173,** 1318–1319 (2018).
144. Posa, I., Carvalho, S., Tavares, J. & Grosso, A. R. A pan-cancer analysis of MYC-PVT1 reveals CNV-unmediated deregulation and poor prognosis in renal carcinoma. *Oncotarget* **7,** 47033–47041 (2016).
145. Tseng, Y.-Y. *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* **512,** 82–86 (2014).
146. Yang, J., Li, C., Mudd, A. & Gu, X. LncRNA PVT1 predicts prognosis and regulates tumor growth in prostate cancer. *Biosci. Biotechnol. Biochem.* **81,** 2301–2306 (2017).
147. Zeng, C. *et al.* Overexpression of the long non-coding RNA PVT1 is correlated with leukemic cell proliferation in acute promyelocytic leukemia. *J. Hematol. Oncol.* **8,** (2015).

148. Zhou, D.-D., Liu, X., Lu, C., Pant, O. P. & Liu, X. Long non-coding RNA PVT1: Emerging biomarker in digestive system cancer. *Cell Prolif.* **50,** (2017).
149. Elmore, S. Apoptosis: a review of programmed cell death. *Toxicol. Pathol.* **35,** 495–516 (2007).
150. Masood, A. *et al.* Downregulation of BCL2 by AT-101 enhances the antileukaemic effect of lenalidomide both by an immune dependant and independent manner. *Br. J. Haematol.* **157,** 59–66 (2012).
151. Lima, R. T., Martins, L. M., Guimarães, J. E., Sambade, C. & Vasconcelos, M. H. Specific downregulation of bcl-2 and xIAP by RNAi enhances the effects of chemotherapeutic agents in MCF-7 human breast cancer cells. *Cancer Gene Ther.* **11,** 309–316 (2004).
152. Tang, S.-W. *et al.* MYC pathway is activated in clear cell renal cell carcinoma and essential for proliferation of clear cell renal cell carcinoma cells. *Cancer Lett.* **273,** 35–43 (2009).
153. Gordan, J. D., Bertout, J. A., Hu, C.-J., Diehl, J. A. & Simon, M. C. HIF-2α Promotes Hypoxic Cell Proliferation by Enhancing c-Myc Transcriptional Activity. *Cancer Cell* **11,** 335–347 (2007).
154. Gordan, J. D. *et al.* HIF-α Effects on c-Myc Distinguish Two Subtypes of Sporadic VHL-Deficient Clear Cell Renal Carcinoma. *Cancer Cell* **14,** 435–446 (2008).
155. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170,** 577–592.e10 (2017).
156. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170,** 564–576.e16 (2017).